

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 04:01:44 ; Search time 25622 Seconds  
(without alignments)  
11676.376 Million cell updates/sec

Title: US-09-242-772-116  
Perfect score: 7313  
Sequence: 1 ggcagcgatacactacaat.....tatgaataaaatctctgtgcc 7313

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_ats.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rnd.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	7313	100.0	7313	9	HSU65002	U65002 Human zinc
2	7312.6	100.0	7313	6	A69511	A69511 Sequence 1
C 3	6520	89.2	142102	9	AC107952	AC107952 Homo sapi
C 4	2821.6	38.6	182102	2	AC097274	AC097274 Mus muscu
C 5	2821.6	38.6	241148	10	AL807387	AL807387 Mouse DNA
C 6	2761.2	37.8	278377	2	AC129839	AC129839 Rattus no
C 7	2623.8	35.9	244955	2	AC123210	AC123210 Rattus no
C 8	1295.6	17.7	1575	10	AF057366	AF057366 Mus muscu
C 9	707.4	9.7	65748	2	AC090187	AC090187 Homo sapi
C 10	651.8	8.9	65748	2	AC090187	AC090187 Homo sapi
C 11	645.4	8.8	49357	2	AC107961	AC107961 Homo sapi
C 12	416	5.7	421	11	G25280	G25280 human STS E
C 13	392.8	5.4	1829	9	BC023655	BC023655 Homo sapi
C 14	392.8	5.4	2429	9	AK026936	AK026936 Homo sapi
C 15	392.8	5.4	4051	9	AF006005	AF006005 Homo sapi
C 16	392.8	5.4	5638	9	D83784	D83784 Human mRNA
C 17	391.2	5.3	5267	10	AB051854	AB051854 Mus muscu
C 18	383.8	5.2	550	9	AF221550	AF221550 Homo sapi
C 19	364.8	5.0	1437	10	AF181262	AF181262 Mus muscu
C 20	359.2	4.9	161478	9	AC025538	AC025538 Homo sapi
C 21	352.8	4.8	145414	9	HSA392M18	HSA392M18 Human DNA
C 22	352.6	4.8	200329	10	AC078911	AC078911 Mus muscu
C 23	352.6	4.8	208614	10	AL807380	AL807380 Mouse DNA
C 24	352	4.8	240510	2	AC134371	AC134371 Rattus no
C 25	351.6	4.8	2334	6	BD005470	BD005470 Nucleic a
C 26	351.6	4.8	2828	9	HSA635	HSA635 Homo sapi
C 27	351.6	4.8	3215	9	AK096331	AK096331 Homo sapi
C 28	351.6	4.8	3663	9	HSA311395	HSA311395 Homo sapi
C 29	351	4.8	240336	2	AC094844	AC094844 Rattus no
C 30	350	4.8	2815	6	AX329760	AX329760 Sequence
C 31	350	4.8	3162	9	HSU72621	HSU72621 Homo sapi
C 32	350	4.8	4833	9	HSU72621	HSU72621 Homo sapi
C 33	344.4	4.7	1874	5	AF186476	AF186476 Danio rer
C 34	333	4.6	376	6	AR274991	AR274991 Sequence
C 35	301.8	4.1	5028	10	RNU72620	RNU72620 Rattus norv
C 36	295.8	4.0	2350	10	AF324471	AF324471 Mus muscu
C 37	294.4	4.0	2771	10	MMZINCFFR	MMZINCFFR M.musculus
C 38	294.4	4.0	2790	6	BD005460	BD005460 Nucleic a
C 39	294.4	4.0	3732	10	MMZINCFFP	MMZINCFFP M.musculus
C 40	294.4	4.0	3975	10	AF147785	AF147785 Mus muscu
C 41	291.4	4.0	236442	2	AC128089	AC128089 Rattus no
C 42	279.2	3.8	2467	9	HSJ468K18	HSJ468K18 Homo sapi
C 43	279.2	3.8	137387	9	HSJ468K18	HSJ468K18 Human DNA
C 44	277.6	3.8	2561	6	A69513	A69513 Sequence 3
C 45	277.6	3.8	2561	9	HSU81992	HSU81992 Homo sapien

# ALIGNMENTS

RESULT 1  
HSU65002  
LOCUS  
DEFINITION Human zinc finger protein PLAG1 mRNA, complete cds.  
ACCESSION U65002  
VERSION U65002.1 GI:1839159  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 7313)  
AUTHORS Kas.K., Voz.M.L., Roijer.E., Astrom,A.K., Meyen,E., Stenman,G. and  
Van de Ven,W.J.  
TITLE Promoter swapping between the genes for a novel zinc finger protein

and beta-catenin in pleiomorphic adenomas with t(3;8) (p21;q12)  
translocations  
Nat. Genet. 15 (2), 170-174 (1997)  
97172974  
PUBMED  
9020842  
REFERENCE  
2 (bases 1 to 7313)  
Kas,K., Voz,M.L., Roijer,E., Meyen,E., Stenman,G. and Van de  
Ven,W.J.M.  
Direct Submission  
Submitted (24-JUL-1996) Lab for Molecular Oncology, Center of Human  
Genetics - K.U.Leuven, Herestraat 49, Leuven 3000, Belgium  
JOURNAL  
Genetics  
Location/Qualifiers  
1..7313  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8q12"  
1..159  
/number=1  
160..264  
/number=2  
265..362  
/number=3  
363..722  
/number=4  
481..1983  
/codon\_start=1  
/product="zinc finger protein PLAG1"  
/protein\_id="AAC50995.1"  
/db\_xref="GI:1839160"  
/translation="MATVIPGDLSEVRDQKVPKGRKGETKPKRKNFFCOLCDKAFN  
SVEKLKVSYSHTGEPIYKCIQDDCTAKFAVSKYKLUORHWATHSPKIKHKNICERMFH  
RKDLKHLNHLHPDKNETFKCEGKNYNTKLGFKRHLALHAATSGDLTKCKVCLQYFE  
STGVLEHLKSHAGKSGGVKKEKHQCHDRFRFYRKDVRHVMVHTGRKDFLOQYC  
AORFGKDLHTRMKSHNOELLKVKTEPVDLPDFTCNVSPVPIKDELLPVMVSLPSE  
LUSKPTNLTQLNLYNTPPQSMQSGSAHQMLITPLGWTCPIDMDTPVSHLSPKY  
PFSSTYALISPKBQPLKGEISYLMELQGVPPSSQDSQASSSKLGLDPOIGSLD  
DGAIDLKSKSISISDPLNTPALDFQNFPIPLNGPPNPLSVLSVLSGMSYSQESAH  
SSVSQPLTQTDQLQDPANTIGLSLHSLSAAFSTSSLSLSTTLPRFHQAFQ"  
723..7313  
/number=5  
2287 a 1327 c 1357 g 2342 t

BASE COUNT  
ORIGIN

Query Match 100.0%; Score 7313; DB 9; Length 7313;  
Best Local Similarity 100.0%; Pred.No. 0;  
Matches 7313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGCAGCGCAGCCGAGGAGGAGTAAAGCCCGCGGTGTAGAG 120  
Db 1 GGCAGCGCAGCCGAGGAGGAGTAAAGCCCGCGGTGTAGAG 120  
QY 61 GCGCGTCCAGCCGAGGAGGAGTAAAGCCCGCGGTGTAGAG 120  
Db 61 GCGCGTCCAGCCGAGGAGGAGTAAAGCCCGCGGTGTAGAG 120  
QY 121 GCGCGGAGCGGCGGAGGAGGAGTAAAGCCCGCGGTGTAGAG 180  
Db 121 GCGCGGAGCGGCGGAGGAGGAGTAAAGCCCGCGGTGTAGAG 180  
QY 181 GCGCGTATTTGGACCCAGAGTCTTCAATCTGTGACGGTCTATTAAAGTTGCCTTG 240  
Db 181 GCGCGTATTTGGACCCAGAGTCTTCAATCTGTGACGGTCTATTAAAGTTGCCTTG 240  
QY 241 CTAGAGTTTGGACAGGCGCTCAGATTGGCCAAAATAGGAAGGATGGATTCCACTCTCT 300  
Db 241 CTAGAGTTTGGACAGGCGCTCAGATTGGCCAAAATAGGAAGGATGGATTCCACTCTCT 300  
QY 301 TCCACGAAGAGTCAATGGGAGTGGCTAAGATCAAGTCTGAGGCTTTTCCATCAGTAAT 360  
Db 301 TCCACGAAGAGTCAATGGGAGTGGCTAAGATCAAGTCTGAGGCTTTTCCATCAGTAAT 360

QY 361 CAGTCCCTTTTGTCTTTTCTTTTACGACCAACATGAACCTTGAGAACCCCTTAAAGCTATA 420  
Db 361 CAGTCCCTTTTGTCTTTTCTTTTACGACCAACATGAACCTTGAGAACCCCTTAAAGCTATA 420  
QY 421 TCATTTAGTGAGTTGGCAGTTCCCAAGTGTCCCAAGAGGAGGCTGTTTGGCTGG 480  
Db 421 TCATTTAGTGAGTTGGCAGTTCCCAAGTGTCCCAAGAGGAGGCTGTTTGGCTGG 480  
QY 481 ATGGCCACTGTCAATCTCTGGTGTATTTGTCAGAGTAAGAGATACCCAGAAAGTCCCTTCA 540  
Db 481 ATGGCCACTGTCAATCTCTGGTGTATTTGTCAGAGTAAGAGATACCCAGAAAGTCCCTTCA 540  
QY 541 GGGAAACGTAGCGTGTGAAACCAACCAAGAAACCTTCTTGGCCAACTGTGTGAC 600  
Db 541 GGGAAACGTAGCGTGTGAAACCAACCAAGAAACCTTCTTGGCCAACTGTGTGAC 600  
QY 601 AAGGCGCTTTAAGTGTGAGAAATTAAGGTTTCACTCTCTACACAGAGAGAGG 660  
Db 601 AAGGCGCTTTAAGTGTGAGAAATTAAGGTTTCACTCTCTACACAGAGAGAGG 660  
QY 661 CCTTCAAGTGCATACCAAGACTGCAACCAAGGCTTCTTCTTAAGTACAAATACAA 720  
Db 661 CCTTCAAGTGCATACCAAGACTGCAACCAAGGCTTCTTCTTAAGTACAAATACAA 720  
QY 721 AGGCACATGGCTACTTCTCTCTGAGAAACCCCAAGTGTAAATTTTGTGAGAAATG 780  
Db 721 AGGCACATGGCTACTTCTCTCTGAGAAACCCCAAGTGTAAATTTTGTGAGAAATG 780  
QY 781 TTTTACCGGAAAGATCATCTGAGAAATCACTTCAACCAAGTGTAAATTTTAAACCTGCTTG 840  
Db 781 TTTTACCGGAAAGATCATCTGAGAAATCACTTCAACCAAGTGTAAATTTTAAACCTGCTTG 840  
QY 841 TTTTACCGGAAAGATCATCTGAGAAATCACTTCAACCAAGTGTAAATTTTAAACCTGCTTG 900  
Db 841 TTTTACCGGAAAGATCATCTGAGAAATCACTTCAACCAAGTGTAAATTTTAAACCTGCTTG 900  
QY 901 GCCTTCGATCGCCGCAACAAAGTGTGACCTCACCTGTAAGAGTATGTTTGAACCTTTTGA 960  
Db 901 GCCTTCGATCGCCGCAACAAAGTGTGACCTCACCTGTAAGAGTATGTTTGAACCTTTTGA 960  
QY 961 AGCAGCGGAGTGTCTTCTGAGGACCTTAAATCTCATGCGGCAAGTCTGCTGGGGTT 1020  
Db 961 AGCAGCGGAGTGTCTTCTGAGGACCTTAAATCTCATGCGGCAAGTCTGCTGGGGTT 1020  
QY 1021 AAGAGAAAGAGCAGGAGTGTGAGAACTTGTATGCGCGGTCTTACACCCGAAAGAGTGC 1080  
Db 1021 AAGAGAAAGAGCAGGAGTGTGAGAACTTGTATGCGCGGTCTTACACCCGAAAGAGTGC 1080  
QY 1081 CGGAGACACATGTTGTCACACTGGAAGAAAGGACTTCTCTGTGATTTGTGACAG 1140  
Db 1081 CGGAGACACATGTTGTCACACTGGAAGAAAGGACTTCTCTGTGATTTGTGACAG 1140  
QY 1141 AGATTTGGGCGAAGAGTCACTGATCGACATATGAAGAGAGTCAATCAAGAGCTT 1200  
Db 1141 AGATTTGGGCGAAGAGTCACTGATCGACATATGAAGAGAGTCAATCAAGAGCTT 1200  
QY 1201 CTGAAGTCAAAACAGAACCCAGTGGATTTCTTGTACCCATTTTACCTGCAATGTCTGTG 1260  
Db 1201 CTGAAGTCAAAACAGAACCCAGTGGATTTCTTGTACCCATTTTACCTGCAATGTCTGTG 1260  
QY 1261 CCTTATAAAGACGAGCTCTTCCGGTGTATGTCCTTACCTTCCAGTGAACCTGTTATCAAG 1320  
Db 1261 CCTTATAAAGACGAGCTCTTCCGGTGTATGTCCTTACCTTCCAGTGAACCTGTTATCAAG 1320  
QY 1321 CCATTCAAAACACTTTGAGATTAACCTCTAACAACCTCAATTTAGTCCATCAGAGC 1380  
Db 1321 CCATTCAAAACACTTTGAGATTAACCTCTAACAACCTCAATTTAGTCCATCAGAGC 1380  
QY 1381 TCGGGATCTGCCACCAAAATGATCAAACTTTTACCTTTGGGATGATGATGATGATGATGAT 1440  
Db 1381 TCGGGATCTGCCACCAAAATGATCAAACTTTTACCTTTGGGATGATGATGATGATGATGAT 1440  
QY 1441 ATGACACTGTTCATCCCTCTCACACCTTTCTTTTCAAAATATCCGTTCTACCTCA 1500

1441 ATGGACATGTTTCATCCCTCTACACCTTTCTTCAATATCCGTTTCAGTTCTACCTCA 1500  
1501 TATGCAATTTCTATTCTGTAAGAAAGACAGCCATTAAAGGGGGAATTTAGAGTTACCTG 1560  
1501 TATGCAATTTCTATTCTGTAAGAAAGACAGCCATTAAAGGGGGAATTTAGAGTTACCTG 1560  
1561 ATGGAGTTACAAGGTGGCGTCCCTCTTCTATCCCAAGATTCTCAAGCATCGTCATCATCT 1620  
1561 ATGGAGTTACAAGGTGGCGTCCCTCTTCTATCCCAAGATTCTCAAGCATCGTCATCATCT 1620  
1621 AAGCTAGGGTTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGAGACCTCTCCCTA 1680  
1621 AAGCTAGGGTTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGAGACCTCTCCCTA 1680  
1681 TCCAAAGGCTCTATCTCCATCAGTACAGCCCTTAAACACACCCAGCATTTGGATTTTCTCAG 1740  
1681 TCCAAAGGCTCTATCTCCATCAGTACAGCCCTTAAACACACCCAGCATTTGGATTTTCTCAG 1740  
1741 TTGTTTAAATTTCAATACCTTTAAATGGTCTCTCCCTATAATCTCTATCAGTGGGAGCCTT 1800  
1741 TTGTTTAAATTTCAATACCTTTAAATGGTCTCTCCCTATAATCTCTATCAGTGGGAGCCTT 1800  
1801 GGAATCAGCTATTTCCAGGAAGACACATTTCTTCTGTTTCCAGCTCCCAACACAACA 1860  
1801 GGAATCAGCTATTTCCAGGAAGACACATTTCTTCTGTTTCCAGCTCCCAACACAACA 1860  
1861 CAGGATCTTCAGGATCTCTCAAAACATATAGGGCTGGGTCTCTGACCTCACTGTGACGA 1920  
1861 CAGGATCTTCAGGATCTCTCAAAACATATAGGGCTGGGTCTCTGACCTCACTGTGACGA 1920  
1921 GCTTTTCAACAGCAGTTTAAAGCAAGTACCAACCTCCCAAGTTTCCATCAAGCTTTTTCAG 1980  
1921 GCTTTTCAACAGCAGTTTAAAGCAAGTACCAACCTCCCAAGTTTCCATCAAGCTTTTTCAG 1980  
1981 TAGGATCTCGGACATGGAATTCATTACAGAAATGATGTAGTGTGCTGCTTGTATTAACAA 2040  
1981 TAGGATCTCGGACATGGAATTCATTACAGAAATGATGTAGTGTGCTGCTTGTATTAACAA 2040  
2041 ATTTTATTTTATGTCCTTCTTAAACAGTATTAATAATTTCTGCTTTTGTATTAACAA 2100  
2041 ATTTTATTTTATGTCCTTCTTAAACAGTATTAATAATTTCTGCTTTTGTATTAACAA 2100  
2101 ATTTTCAATTAAGCCAGTATATAAATAAGAACTAGCTTTTAACTGAGCTTTGGAACCAATT 2160  
2101 ATTTTCAATTAAGCCAGTATATAAATAAGAACTAGCTTTTAACTGAGCTTTGGAACCAATT 2160  
2161 GTGTTCAAGTTAAGTTTACTGGGTATTTTGTCTGATTCATGCAATTTGTACATTTTA 2220  
2161 GTGTTCAAGTTAAGTTTACTGGGTATTTTGTCTGATTCATGCAATTTGTACATTTTA 2220  
2221 AGACTTTTTTTTTTCCATATAGGAAGCCATTATTAGTAGTAACCTTTTACAAATCCCA 2280  
2221 AGACTTTTTTTTTTCCATATAGGAAGCCATTATTAGTAGTAACCTTTTACAAATCCCA 2280  
2281 TTTTCAAAATTAAGTTTATAGATCTTAAATTTTCAATTTTGTCTAATAAGAGTTTGT 2340  
2281 TTTTCAAAATTAAGTTTATAGATCTTAAATTTTCAATTTTGTCTAATAAGAGTTTGT 2340  
2341 TTTTGAATCTGGCTCATTAATAAATTTAGCAATGAGATGATGATTAATAAGAGTTTGT 2400  
2341 TTTTGAATCTGGCTCATTAATAAATTTAGCAATGAGATGATGATTAATAAGAGTTTGT 2400  
2401 TGAATAACTCAAGGTTTAAATTTTCTTACTAGTCTTCAATGATTAATAACTCAAGTGC 2460  
2401 TGAATAACTCAAGGTTTAAATTTTCTTACTAGTCTTCAATGATTAATAACTCAAGTGC 2460  
2461 TTCAATGAATTAAGAGTCCAGTTTCGGAAGTAAATAAATTTTGTGTAGATACACCATAA 2520  
2461 TTCAATGAATTAAGAGTCCAGTTTCGGAAGTAAATAAATTTTGTGTAGATACACCATAA 2520  
2521 TTTTCAGATCAGTATATTCTGAAGACTCTCTGTCTGGCTTAAATATTTTGGCATCTTTTA 2580  
2521 TTTTCAGATCAGTATATTCTGAAGACTCTCTGTCTGGCTTAAATATTTTGGCATCTTTTA 2580

2521 TTTTCAGATCAGTATATTCTGAAGACTCTCTGTCTGGCTTAAATATTTTGGCATCTTTTA 2580  
2581 TTTTCAGATCAGTATATTCTGAAGACTCTCTGTCTGGCTTAAATATTTTGGCATCTTTTA 2580  
2581 TTTTCAGATCAGTATATTCTGAAGACTCTCTGTCTGGCTTAAATATTTTGGCATCTTTTA 2580  
2641 AGACTCTGTAGGTGATGGCATTTTCGTGTCACTTAAATAGTTGGCCCTAAATTAGTACA 2700  
2641 AGACTCTGTAGGTGATGGCATTTTCGTGTCACTTAAATAGTTGGCCCTAAATTAGTACA 2700  
2701 CAGGATATTTTGTCTGTTTTCATCTTCTTAAATAGTTGGCCCTAAATTAGTACA 2760  
2701 CAGGATATTTTGTCTGTTTTCATCTTCTTAAATAGTTGGCCCTAAATTAGTACA 2760  
2761 ATAGTGTATGGCATTTGGGTCTTTCAGAGTCGATATATAGGTAGATCTCTTTAGTCTTTTC 2820  
2761 ATAGTGTATGGCATTTGGGTCTTTCAGAGTCGATATATAGGTAGATCTCTTTAGTCTTTTC 2820  
2821 CACCTTTTCAATCCCAAGGGTGGGTCAAGTGCAGCCAGCAATTTATTTTCAITGTTGGCC 2880  
2821 CACCTTTTCAATCCCAAGGGTGGGTCAAGTGCAGCCAGCAATTTATTTTCAITGTTGGCC 2880  
2881 CACGGTTAGTCCATATCTAGAGCCATCTGGAATCTGCAGCCATGAGGTGTGTTTATCC 2940  
2881 CACGGTTAGTCCATATCTAGAGCCATCTGGAATCTGCAGCCATGAGGTGTGTTTATCC 2940  
2941 ACAGTGGATTTAGTCTGAGCTCTGTTGGTGCAGACTTCTTAAGCAGGAATAGACGTGAA 3000  
2941 ACAGTGGATTTAGTCTGAGCTCTGTTGGTGCAGACTTCTTAAGCAGGAATAGACGTGAA 3000  
3001 GCACATGTTTACATTTGGGAACTTGTAGGATCATGGCCCTGTAGCCAGGGTTAAAA 3060  
3001 GCACATGTTTACATTTGGGAACTTGTAGGATCATGGCCCTGTAGCCAGGGTTAAAA 3060  
3061 ACTGACTTTTTTAGAAGTAAAGTAAAGCAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3120  
3061 ACTGACTTTTTTAGAAGTAAAGTAAAGCAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3120  
3121 ATATGTTTTTCTTTTCCCTTAAAGAAATCAAAAGCAGAAAAACAAACAGTCTACTCCG 3180  
3121 ATATGTTTTTCTTTTCCCTTAAAGAAATCAAAAGCAGAAAAACAAACAGTCTACTCCG 3180  
3181 ATGTTATCTTTCTGATTCATGTAATCCATCTTCTTCCCTGCAATATTTTGGATGAGAA 3240  
3181 ATGTTATCTTTCTGATTCATGTAATCCATCTTCTTCCCTGCAATATTTTGGATGAGAA 3240  
3241 TTGAAGTTAAATGCAATTAAGAACTTACCTGATGAACTACCAAGTAAAGTAAAGTAAAGTAAAG 3300  
3241 TTGAAGTTAAATGCAATTAAGAACTTACCTGATGAACTACCAAGTAAAGTAAAGTAAAGTAAAG 3300  
3301 AAATATATACAGTAAAAATCCCACTTTTTCATGATCTCTGGGAAATGATAGGATTTGCAA 3360  
3301 AAATATATACAGTAAAAATCCCACTTTTTCATGATCTCTGGGAAATGATAGGATTTGCAA 3360  
3361 ATAGTTTGAAGTTTGTAGAGGTAAACAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3420  
3361 ATAGTTTGAAGTTTGTAGAGGTAAACAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3420  
3421 GAAATTAACAAATTCAGAAATATTTATATTTTCACTGAAATATATATATATATATATATATATATAT 3480  
3421 GAAATTAACAAATTCAGAAATATTTATATTTTCACTGAAATATATATATATATATATATATATATAT 3480  
3481 AGACAACCAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3540  
3481 AGACAACCAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3540  
3541 GACATTAACCTTAAGAAACATATCTTAAAGCACTTGTAAACACTTATTTTAGGACCTT 3600  
3541 GACATTAACCTTAAGAAACATATCTTAAAGCACTTGTAAACACTTATTTTAGGACCTT 3600  
3601 TCCGTTGGGATCGGGGCAAGGGTGGGAGGTTTTTAGAAGAGTATATATCTCTTTAAAAA 3660  
3601 TCCGTTGGGATCGGGGCAAGGGTGGGAGGTTTTTAGAAGAGTATATATCTCTTTAAAAA 3660





QY 61 GCCGGTCCAGCCGCAATATAGAAAAAATATTAGAAATTCGCGCGCGGTGTAGAG 120  
DB 61 GCCGGTCCAGCCGCAATATAGAAAAAATATTAGAAATTCGCGCGCGGTGTAGAG 120  
QY 121 GCGGGGACGCGCGGAGGAGATGTTAAAGCCCGCGGTGCGCTTTGGTGTGCGCTT 180  
DB 121 GCGGGGACGCGCGGAGGAGATGTTAAAGCCCGCGGTGCGCTTTGGTGTGCGCTT 180  
QY 181 GCGCGTATTGGACCCAGAAATGCTTCATCTGTGACGGTCTATTATTAAGTTGCGCTG 240  
DB 181 GCGCGTATTGGACCCAGAAATGCTTCATCTGTGACGGTCTATTATTAAGTTGCGCTG 240  
QY 241 CTAGAGTTTGGACGCGGCTCAGATTGGCAAAATGGGAAGATTGGATTCCACTCTCT 300  
DB 241 CTAGAGTTTGGACGCGGCTCAGATTGGCAAAATGGGAAGATTGGATTCCACTCTCT 300  
QY 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAACTGTGAGGCTTTTCCATCAGTAAT 360  
DB 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAACTGTGAGGCTTTTCCATCAGTAAT 360  
QY 361 CAGTCCCTTTTGGTCTTTTACGACCAATGAAACTTGAAGCCACCTAAGCTATA 420  
DB 361 CAGTCCCTTTTGGTCTTTTACGACCAATGAAACTTGAAGCCACCTAAGCTATA 420  
QY 421 TCATTTAGTGGAGTTGGGCAAGTTCCCAAGTGTCCAAAGAGGCGCTTTAGGCTGCG 480  
DB 421 TCATTTAGTGGAGTTGGGCAAGTTCCCAAGTGTCCAAAGAGGCGCTTTAGGCTGCG 480  
QY 481 ATGCGCACTGTCAATTCCTGTGTGTTGTGAGAGTAAAGATACCCAGAAAGTCCCTTCA 540  
DB 481 ATGCGCACTGTCAATTCCTGTGTGTTGTGAGAGTAAAGATACCCAGAAAGTCCCTTCA 540  
QY 541 GGGAAAGTAAAGCGTGTGAAACCAACCAAGAAACCTTCCCTGCCAAGTGTGAC 600  
DB 541 GGGAAAGTAAAGCGTGTGAAACCAACCAAGAAACCTTCCCTGCCAAGTGTGAC 600  
QY 601 AAGGCTTTTAAACAGTGTGAGAAATTAAGGTTTCACTTCTACTCTCACAGAGAGAGG 660  
DB 601 AAGGCTTTTAAACAGTGTGAGAAATTAAGGTTTCACTTCTACTCTCACAGAGAGAGG 660  
QY 661 CCCTACAAGTGCATACAAAGACTGCAACCAAGGCTTTGTTTCTAAGTACAAATTAACA 720  
DB 661 CCCTACAAGTGCATACAAAGACTGCAACCAAGGCTTTGTTTCTAAGTACAAATTAACA 720  
QY 721 AGGCATATGGTACTCATCTCTGAGAAACCAAGTGTAAATTTGTGAGAAATG 780  
DB 721 AGGCATATGGTACTCATCTCTGAGAAACCAAGTGTAAATTTGTGAGAAATG 780  
QY 781 TTTTCCCGGAAAGATCATCTGAAGATCACTCCATACACAGCACCCCTAACAAAGAGAG 840  
DB 781 TTTTCCCGGAAAGATCATCTGAAGATCACTCCATACACAGCACCCCTAACAAAGAGAG 840  
QY 841 TTTAAGTGCAGAAATGTGGCAAGACTTAAATACCAAGTGTGAATTTAAGCTCCTTG 900  
DB 841 TTTAAGTGCAGAAATGTGGCAAGACTTAAATACCAAGTGTGAATTTAAGCTCCTTG 900  
QY 901 GCCTTGATGCGCAACCAAGTGTGACCTCACTGTAAGTATGTTGCAAACTTTTGA 960  
DB 901 GCCTTGATGCGCAACCAAGTGTGACCTCACTGTAAGTATGTTGCAAACTTTTGA 960  
QY 961 AGCAGGGAGTGTCTCGAGCACCTTAAATCTCATGAGGCAAGTCTGTGTTGGGGTT 1020  
DB 961 AGCAGGGAGTGTCTCGAGCACCTTAAATCTCATGAGGCAAGTCTGTGTTGGGGTT 1020  
QY 1021 AAAGAAAAAGACCAAGTGGCAACATTTGATGCGCGGTCTACACCCGAAAGATGTC 1080  
DB 1021 AAAGAAAAAGACCAAGTGGCAACATTTGATGCGCGGTCTACACCCGAAAGATGTC 1080  
QY 1081 CGGAGACATGTTGGTGCACACTGGAAGAAAGACTTCTCTGTGAGTATTGTGCACAG 1140  
DB 1081 CGGAGACATGTTGGTGCACACTGGAAGAAAGACTTCTCTGTGAGTATTGTGCACAG 1140  
QY 1141 AGATTGGCGGAAAGGATCACCTGACTCGACATATGAAGAAGAGTCACAATCAAGAGCTT 1200

DB 1141 AGATTGGCGGAAAGGATCACCTGACTCGACATATGAAGAAGAGTCACAATCAAGAGCTT 1200  
QY 1201 CTGAAGGTCAAAACAGAACCAAGTGGATTTCTTTGACCCATTTACCTGCAATGTGTG 1260  
DB 1201 CTGAAGGTCAAAACAGAACCAAGTGGATTTCTTTGACCCATTTACCTGCAATGTGTG 1260  
QY 1261 CCTATAAAGACAGAGTCTCTCCGGTGATGCTTACCTTCCAGTGAATGTATCAAG 1320  
DB 1261 CCTATAAAGACAGAGTCTCTCCGGTGATGCTTACCTTCCAGTGAATGTATCAAG 1320  
QY 1321 CAATTCAAAAACACTTTGAGTTAAACCTCTAACACATCCATTTTCCAGTCCATGAGAGC 1380  
DB 1321 CAATTCAAAAACACTTTGAGTTAAACCTCTAACACATCCATTTTCCAGTCCATGAGAGC 1380  
QY 1381 TCGGATCTGCCACCACCAATGATCAAACTTTACCTTTGGGAATGATGCCCAATAGAT 1440  
DB 1381 TCGGATCTGCCACCACCAATGATCAAACTTTACCTTTGGGAATGATGCCCAATAGAT 1440  
QY 1441 ATGACACTGTTTCATCCCTCTCACACCTTTCTTTCAAAATATCCGTTTCACTCTCA 1500  
DB 1441 ATGACACTGTTTCATCCCTCTCACACCTTTCTTTCAAAATATCCGTTTCACTCTCA 1500  
QY 1501 TATGCAATTTCTATTCTGAAAAGAACAGCCATTAAGGGGAAATGAGAGTTACCTG 1560  
DB 1501 TATGCAATTTCTATTCTGAAAAGAACAGCCATTAAGGGGAAATGAGAGTTACCTG 1560  
QY 1561 ATGAGTTTCAAGGTGGCGTCCCTCTTCAATCCCAAGATTTCTCAAGCATCGTCATCT 1620  
DB 1561 ATGAGTTTCAAGGTGGCGTCCCTCTTCAATCCCAAGATTTCTCAAGCATCGTCATCT 1620  
QY 1621 AAGCTAGGTTGATCTCTCAGATTGGTCCCTAGATGATGGTCCAGAGACCTCTCCCTA 1680  
DB 1621 AAGCTAGGTTGATCTCTCAGATTGGTCCCTAGATGATGGTCCAGAGACCTCTCCCTA 1680  
QY 1681 TCCAAAAGCTTATCTCCATCAGTGCACCCCTTAAACACACAGCATTGGATTTTCTCAG 1740  
DB 1681 TCCAAAAGCTTATCTCCATCAGTGCACCCCTTAAACACACAGCATTGGATTTTCTCAG 1740  
QY 1741 TTGTTTAAATTTATACCTTTAAATGGTCTCCCTATATATCTCTATCAGTGGGAGCCTT 1800  
DB 1741 TTGTTTAAATTTATACCTTTAAATGGTCTCCCTATATATCTCTATCAGTGGGAGCCTT 1800  
QY 1801 GGAATGAGCTATTCCCAAGGAAGACATCTTCTCTGTTTCCAGCTCCCAACACAAACA 1860  
DB 1801 GGAATGAGCTATTCCCAAGGAAGACATCTTCTCTGTTTCCCAAGCTCCCAACACAAACA 1860  
QY 1861 CAGGATCTTCCAGGATCTGCAAAACATATAGGGCTTGGGCTCTGCACTCACTCTCAGCA 1920  
DB 1861 CAGGATCTTCCAGGATCTGCAAAACATATAGGGCTTGGGCTCTGCACTCACTCTCAGCA 1920  
QY 1921 GCTTTCCAGGAGTATTAAGCAAGTACCACCCCTCCAGCTTTCCATCAAGCTTTTCAAG 1980  
DB 1921 GCTTTCCAGGAGTATTAAGCAAGTACCACCCCTCCAGCTTTCCATCAAGCTTTTCAAG 1980  
QY 1981 TAGGATTTCTGGGACATGATTCATTAAGAAATGATGTGTAGCTGTGCGCTTACATGACC 2040  
DB 1981 TAGGATTTCTGGGACATGATTCATTAAGAAATGATGTGTAGCTGTGCGCTTACATGACC 2040  
QY 2041 ATTTTATTTTGTAGTGCCTTCTTTAAACAGTATATAAATTTCTGTTTGTATATAACA 2100  
DB 2041 ATTTTATTTTGTAGTGCCTTCTTTAAACAGTATATAAATTTCTGTTTGTATATAACA 2100  
QY 2101 ATTTTCTAATAGCCAGTATATAAATAGAACTTAGCTTTTAAACTGAGCTTTGGAACCAT 2160  
DB 2101 ATTTTCTAATAGCCAGTATATAAATAGAACTTAGCTTTTAAACTGAGCTTTGGAACCAT 2160  
QY 2161 GTGTTCAGTTAAGTTTACCTGGGTATTTTGTGCTGATTCATGCTCAATTTGACATTTTA 2220  
DB 2161 GTGTTCAGTTAAGTTTACCTGGGTATTTTGTGCTGATTTCTGCTGATTTGACATTTTA 2220  
QY 2221 AGACTTTTCTTTTCCATATAGGAAGCCATTTATAGTAGTAAACTTTTACAAATCCCA 2280

Db 2221 AGACCTTTTTCATATAGGAAGCCATTATTAGTAGTAACCTTTTACAAATCCCA 2280  
Qy 2281 TTTTCAAAATTAATTTTAGATCTTAAATTTTCAATTTTGTCTAAATAACAGTGGCTCTACC 2340  
Db 2281 TTTTCAAAATTAATTTTAGATCTTAAATTTTCAATTTTGTCTAAATAACAGTGGCTCTACC 2340  
Qy 2341 TTTTGACATCTGGCTCAATTAATAATTTAGCAATAGAAATGTAATGTAATTAATAAGTTTG 2400  
Db 2341 TTTTGACATCTGGCTCAATTAATAATTTAGCAATAGAAATGTAATGTAATTAATAAGTTTG 2400  
Qy 2401 TGAATAACTCAAGGGTTTAAATTTTCTTACTAGCTTCTAAATGCAATTAATCAAGTGC 2460  
Db 2401 TGAATAACTCAAGGGTTTAAATTTTCTTACTAGCTTCTAAATGCAATTAATCAAGTGC 2460  
Qy 2461 TTCAATGAATTAAGAGTCCAGTTTCGGAAGATATAAATGTTTGTAGATACACCAATAA 2520  
Db 2461 TTCAATGAATTAAGAGTCCAGTTTCGGAAGATATAAATGTTTGTAGATACACCAATAA 2520  
Qy 2521 TTTTCAGATCAGTATATTCTGAAGCTCTCTGTGTCTGGCTAAATATTTTGCCATCTTTA 2580  
Db 2521 TTTTCAGATCAGTATATTCTGAAGCTCTCTGTGTCTGGCTAAATATTTTGCCATCTTTA 2580  
Qy 2581 TTATGAGCCCTTAAAGGAACCAACCCCTAAACCAAGCAATCAGTATTTAGCAAAAAG 2640  
Db 2581 TTATGAGCCCTTAAAGGAACCAACCCCTAAACCAAGCAATCAGTATTTAGCAAAAAG 2640  
Qy 2641 AGACTCTGTTAGTGATGATGGCAATTTCTGTCTCACTTAATAGTTGGCCCTAAATTTAGTACA 2700  
Db 2641 AGACTCTGTTAGTGATGATGGCAATTTCTGTCTCACTTAATAGTTGGCCCTAAATTTAGTACA 2700  
Qy 2701 CAGGATATTTTGTGTTTTCATCTCTTTTAAACATGCTATCTTTTCAATTTAATAAGTA 2760  
Db 2701 CAGGATATTTTGTGTTTTCATCTCTTTTAAACATGCTATCTTTTCAATTTAATAAGTA 2760  
Qy 2761 ATAGTCTATGGCAATGGGTCCTTACAGTCTGATATATAGTAGATCTCTTTAGTCTTTTC 2820  
Db 2761 ATAGTCTATGGCAATGGGTCCTTACAGTCTGATATATAGTAGATCTCTTTAGTCTTTTC 2820  
Qy 2821 CACCTTTTCATCCAAAGGGTGGTCAAGTGCAGCCAGCAATTTATTTTCAATTTGTTGGC 2880  
Db 2821 CACCTTTTCATCCAAAGGGTGGTCAAGTGCAGCCAGCAATTTATTTTCAATTTGTTGGC 2880  
Qy 2881 CACGGTTAGTCCATATCTAGAGCCATTTGTGAATCTGACCCATAGGTTGTTATCCC 2940  
Db 2881 CACGGTTAGTCCATATCTAGAGCCATTTGTGAATCTGACCCATAGGTTGTTATCCC 2940  
Qy 2941 ACAGTGGATTGACTCAGCCCTCTGTGGTGCACAGCTTCTTAAGCAGCAATAGACGTGAA 3000  
Db 2941 ACAGTGGATTGACTCAGCCCTCTGTGGTGCACAGCTTCTTAAGCAGCAATAGACGTGAA 3000  
Qy 3001 GCACATGGTTACATTTGGGAATCTGTGTAGGATCATGGCCCTGTAGCCAGGGTTAAAA 3060  
Db 3001 GCACATGGTTACATTTGGGAATCTGTGTAGGATCATGGCCCTGTAGCCAGGGTTAAAA 3060  
Qy 3061 ACTGACATTTTGAAGTAAGTAAGATAGAGCTTATATCAATTTCTGTGTAATTTG 3120  
Db 3061 ACTGACATTTTGAAGTAAGTAAGATAGAGCTTATATCAATTTCTGTGTAATTTG 3120  
Qy 3121 ATATGTTTTTCTTCCCTTAAGCAATCAAGCAGCAAAACCAACAGTCTCTCCG 3180  
Db 3121 ATATGTTTTTCTTCCCTTAAGCAATCAAGCAGCAAAACCAACAGTCTCTCTCCG 3180  
Qy 3181 ATGTTATCTTTCTGATCGAATCCATCTTTCTTGCATATTTTGGATGAGAT 3240  
Db 3181 ATGTTATCTTTCTGATCGAATCCATCTTTCTTGCATATTTTGGATGAGAT 3240  
Qy 3241 TTGAAGTTTAATGCAATTAAGAACTACCTGATGAATACCAAAAGTTTAAAGTACTAG 3300  
Db 3241 TTGAAGTTTAATGCAATTAAGAACTACCTGATGAATACCAAAAGTTTAAAGTACTAG 3300  
Qy 3301 AAATATATACAGTAATTAATCCATTTTCAATGCTCTCTGGGAAATGATAGGATATTGCAA 3360  
Db 3301 AAATATATACAGTAATTAATCCATTTTCAATGCTCTCTGGGAAATGATAGGATATTGCAA 3360

Qy 3361 ATAAAGTTGAGTTTGTAGAGGGTAACAAGTAAGTAACAAACAACTATCTTTGGTTAAACAT 3420  
Db 3361 ATAAAGTTGAGTTTGTAGAGGGTAACAAGTAAGTAACAAACAACTATCTTTGGTTAAACAT 3420  
Qy 3421 GAAATAACAATTCAGAAATATATATATTCACATCAATTAATATAGGCTTTTCTTCAAT 3480  
Db 3421 GAAATAACAATTCAGAAATATATATATTCACATCAATTAATATAGGCTTTTCTTCAAT 3480  
Qy 3481 AGACAACAACATATCTTTAAAGGTCTAAATTAATATATTTTCTAAAGGTCAGTTGG 3540  
Db 3481 AGACAACAACATATCTTTAAAGGTCTAAATTAATATATTTTCTAAAGGTCAGTTGG 3540  
Qy 3541 GACATTAACCTTAAGAAACATATCTTTTAAGCACTTTGTTAACACCTTATTTTAGGACCTT 3600  
Db 3541 GACATTAACCTTAAGAAACATATCTTTTAAGCACTTTGTTAACACCTTATTTTAGGACCTT 3600  
Qy 3601 TCCGTTGGGGATGGGCAAGGGTGGGAGGTGTTTAAAGAGTATATATCTTTTAAAAA 3660  
Db 3601 TCCGTTGGGGATGGGCAAGGGTGGGAGGTGTTTAAAGAGTATATATCTTTTAAAAA 3660  
Qy 3661 AAAACAGAAAGAAAAATTTTCTGAGCACTCATTTAGCCCTATATGGAACCTTTCTTCCCT 3720  
Db 3661 AAAACAGAAAGAAAAATTTTCTGAGCACTCATTTAGCCCTATATGGAACCTTTCTTCCCT 3720  
Qy 3721 TTTGTAGGCCAGTTTATCTGAGCACTTTGCAATTTTACCAAGAAATTTCTAAAAATGAGT 3780  
Db 3721 TTTGTAGGCCAGTTTATCTGAGCACTTTGCAATTTTACCAAGAAATTTCTAAAAATGAGT 3780  
Qy 3781 GCAGATTAATGAATAATAACATTTTAAATATTTGGGAGTAGTATATTTGTTGAGA 3840  
Db 3781 GCAGATTAATGAATAATAACATTTTAAATATTTGGGAGTAGTATATTTGTTGAGA 3840  
Qy 3841 AATGTAATTTGAATAATGTAATGGGGCTTCAATATATATATATATATATATACACACAC 3900  
Db 3841 AATGTAATTTGAATAATGTAATGGGGCTTCAATATATATATATATATATACACACAC 3900  
Qy 3901 ACACATGACACACATACCGCACTTCAATAGAAATCAAAAGTTGCTCTGAAGGAGCTTTGG 3960  
Db 3901 ACACATGACACACATACCGCACTTCAATAGAAATCAAAAGTTGCTCTGAAGGAGCTTTGG 3960  
Qy 3961 CTCCTGATATTTTATCATGCTCTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 4020  
Db 3961 CTCCTGATATTTTATCATGCTCTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 4020  
Qy 4021 TTATGATTTTAAATTTTATTTATGAAATTTAATTTTAAATTTTAAATTTTAAATTTTAAAT 4080  
Db 4021 TTATGATTTTAAATTTTATTTATGAAATTTAATTTTAAATTTTAAATTTTAAATTTTAAAT 4080  
Qy 4081 ATTTAAATTTATATATTTTAAAGGAAGTACATTTTAAATTTTAAATTTTAAATTTTAAAT 4140  
Db 4081 ATTTAAATTTATATATTTTAAAGGAAGTACATTTTAAATTTTAAATTTTAAATTTTAAAT 4140  
Qy 4141 TTTTGAATTTAGGAGAAATTTTCTGAAAGTAGGCTTTTCTGCTCTCTCTCTCTCTCTCTCT 4200  
Db 4141 TTTTGAATTTAGGAGAAATTTTCTGAAAGTAGGCTTTTCTGCTCTCTCTCTCTCTCTCTCT 4200  
Qy 4201 CTTTGTAGTTTCTATGAAACAGATTTTCTTACCTTAAATCTTTTAAATTTTAAATTTTAAAT 4260  
Db 4201 CTTTGTAGTTTCTATGAAACAGATTTTCTTACCTTAAATCTTTTAAATTTTAAATTTTAAAT 4260  
Qy 4261 ATATGCTTTTAAATCAACATATAAAGGAAAAAATTTGGTGACAGACACAAATAGAAAC 4320  
Db 4261 ATATGCTTTTAAATCAACATATAAAGGAAAAAATTTGGTGACAGACACAAATAGAAAC 4320  
Qy 4321 CTATTTTAAATAGAAATCAAAATAGCAAGTGTGGAAGCACTTCTTTTCTGTTTAAA 4380  
Db 4321 CTATTTTAAATAGAAATCAAAATAGCAAGTGTGGAAGCACTTCTTTTCTGTTTAAA 4380  
Qy 4381 ATGACTTTAAGAGTCAATCAAAATAGTGAACCTGAGACATTTGGCCCTTAGTAGGCTGATTC 4440  
Db 4381 ATGACTTTAAGAGTCAATCAAAATAGTGAACCTGAGACATTTGGCCCTTAGTAGGCTGATTC 4440







```

Db      6601 CATGTGCACTGCAAAAGGAGATCTCGACATGCAAAATGTTACCGAACAACCCAGCTT 6660
QY      6661 TTGTCCACAAGGTGACTGTAACATCAAGATGGAAGTGGCTTTATATAGGGTGTGGACT 6720
Db      6661 TTGTCCACAAGGTGACTGTAACATCAAGATGGAAGTGGCTTTATATAGGGTGTGGACT 6720
QY      6721 GAAGAACATGCTGTATGTTACTAACAAGCCCTTTGAATTTTAAACAAAACATGGGAATCCATT 6780
Db      6721 GAAGAACATGCTGTATGTTACTAACAAGCCCTTTGAATTTTAAACAAAACATGGGAATCCATT 6780
QY      6781 AGGAACGATGTCATCATCACTGACATCAAGTGAAGTGGCTGCTGAAATTTGATTTTGTAGCT 6840
Db      6781 AGGAACGATGTCATCATCACTGACATCAAGTGAAGTGGCTGCTGAAATTTGATTTTGTAGCT 6840
QY      6841 AATGAAAAAGTGTGGCACTAGTACTCTAATAATGTTCTAATGATATAAGTGTGGTGTGAGTCAA 6900
Db      6841 AATGAAAAAGTGTGGCACTAGTACTCTAATAATGTTCTAATGATATAAGTGTGGTGTGAGTCAA 6900
QY      6901 AATGAAAAAGAAAAAATCTGCATTCAGCCCGCAATTTTGTATATTTTATTTGCAATTTAA 6960
Db      6901 AATGAAAAAGAAAAAATCTGCATTCAGCCCGCAATTTTGTATATTTTATTTGCAATTTAA 6960
QY      6961 AATTCCTATTCTGTAATATTTGGGAATCAAGTGGCTTATCATGTATATCGTGTACTTAA 7020
Db      6961 AATTCCTATTCTGTAATATTTGGGAATCAAGTGGCTTATCATGTATATCGTGTACTTAA 7020
QY      7021 ATGTATTACAAAACACTGTTGTTGTTATTTGTAATAATATAGACAAAGATCATATTTTGT 7080
Db      7021 ATGTATTACAAAACACTGTTGTTGTTATTTGTAATAATATAGACAAAGATCATATTTTGT 7080
QY      7081 GTGTGTATAAGCTCTGTAATAATAGCAATACATTTATGAAGCTGCAGTGATACATATTT 7140
Db      7081 GTGTGTATAAGCTCTGTAATAATAGCAATACATTTATGAAGCTGCAGTGATACATATTT 7140
QY      7141 AAACATTCACATCCAAAGACAGACTATTTATGTCGCATATACAGATTTAAATATTA 7200
Db      7141 AAACATTCACATCCAAAGACAGACTATTTATGTCGCATATACAGATTTAAATATTA 7200
QY      7201 ATTTCCTGTAATAATAATAGTACTGAGCTTCTGTGGCTACAGTGTATGTTGCTG 7260
Db      7201 ATTTCCTGTAATAATAATAGTACTGAGCTTCTGTGGCTACAGTGTATGTTGCTG 7260
QY      7261 TGTAAGAATAAGATATGTAATTCACAAAATATATGTAATAAATCTGCTGCC 7313
Db      7261 TGTAAGAATAAGATATGTAATTCACAAAATATATGTAATAAATCTGCTGCC 7313

RESULT 3
AC107952/c
LOCUS      AC107952 Homo sapiens chromosome 8, clone RP11-140I16, complete sequence.
DEFINITION AC107952
ACCESSION  AC107952
VERSION    AC107952.5 GI:21930261
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 142102)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
TITLE      Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
JOURNAL    Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
REFERENCE  Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
AUTHORS    Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
TITLE      Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
JOURNAL    Hagsos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
REFERENCE  Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
AUTHORS    Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,

```

```

Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Rise,C., Rogov,P.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 142102)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 142102)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cooke,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 23, 2002 this sequence version replaced gi:21699524.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24490
Center clone name: 140_I_16
----- Location/Qualifiers

```

FEATURES



Db 30559 TGCCATATAAAGACGAGCTCCTTCGGTGATGTCTTACCTCCAGTGAACCTGTTATCAA 30500  
QY 1319 AGCCATTACAACACCTTTGCACTTAAACCTCTAGAACACCTCATTTAGTCCATGCAGA 1378  
Db 30499 AGCCATTACAACACCTTTGCACTTAAACCTCTAGAACACCTCATTTAGTCCATGCAGA 30440  
QY 1379 GCTCGGAGTCTGCCACCAAAATGATCACAACTTTTACCTTTGGGAATGACATGCCCAATAG 1438  
Db 30439 GCTCGGAGTCTGCCACCAAAATGATCACAACTTTTACCTTTGGGAATGACATGCCCAATAG 30380  
QY 1439 ATATGGACACTGTTGATCCCTCTGACCACTTTCTTTTCAAATGATCGTTCTACCT 1498  
Db 30379 ATATGGACACTGTTGATCCCTCTGACCACTTTCTTTTCAAATGATCGTTCTACCT 30320  
QY 1499 CATATGCAATTTCTATTTCTGAAAAGACAGCCATTAAAGGGGGAATTTGAGAGTTACC 1558  
Db 30319 CATATGCAATTTCTATTTCTGAAAAGACAGCCATTAAAGGGGGAATTTGAGAGTTACC 30260  
QY 1559 TGATGGAGTTACAAGTGGCGTGCCTCTTCTATCCCAAGATTTCTCAAGCATCGTCAATCAT 1618  
Db 30259 TGATGGAGTTACAAGTGGCGTGCCTCTTCTATCCCAAGATTTCTCAAGCATCGTCAATCAT 30200  
QY 1619 CTAAGCTAGGTTGATCCTCAGATGGTCCCTAGATGATGGTGAGAGACCTCTCC 1678  
Db 30199 CTAAGCTAGGTTGATCCTCAGATGGTCCCTAGATGATGGTGAGAGACCTCTCC 30140  
QY 1679 TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACACAGCATTTGGATTTTCTC 1738  
Db 30139 TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACACAGCATTTGGATTTTCTC 30080  
QY 1739 AGTTGTTTAAATTCATACCTTTTAAATGGTCTCCCTCTATAATCCTCTATCAGTGGGAGCC 1798  
Db 30079 AGTTGTTTAAATTCATACCTTTTAAATGGTCTCCCTCTATAATCCTCTATCAGTGGGAGCC 30020  
QY 1799 TTGGAATGAGCTATTTCCAGGAGAGACACATCTCTGTTTCCAGCTCCCCACACAAA 1858  
Db 30019 TTGGAATGAGCTATTTCCAGGAGAGACACATCTCTGTTTCCAGCTCCCCACACAAA 29960  
QY 1859 CACAGGATCTTCAGGATCCTGCAACACTATAGGCTTTGGGTCTCTGCACCTCACTGTCTAG 1918  
Db 29959 CACAGGATCTTCAGGATCCTGCAACACTATAGGCTTTGGGTCTCTGCACCTCACTGTCTAG 29900  
QY 1919 CAGCTTCCAGCAGAGTTTAAAGCACAGTACACCCCTCCAGCTTTCCATCAAGCTTTTC 1978  
Db 29899 CAGCTTCCAGCAGAGTTTAAAGCACAGTACACCCCTCCAGCTTTCCATCAAGCTTTTC 29840  
QY 1979 AGTAGGATCTGGGACATGGATTTATTAAGAAATGATGTAGCTGTGCCCTAGATGA 2038  
Db 29839 AGTAGGATCTGGGACATGGATTTATTAAGAAATGATGTAGCTGTGCCCTAGATGA 29780  
QY 2039 CCATTTTATTTTATGTCCTACTTTTAAACAGTATAAAATTTCTGCTTTGTATATATAC 2098  
Db 29779 CCATTTTATTTTATGTCCTACTTTTAAACAGTATAAAATTTCTGCTTTGTATATATAC 29720  
QY 2099 AAATTTTCATTAAGCCAGTATAAATAGAACTAGCTTTTAAACTGACCTTTGGAACCAT 2158  
Db 29719 AAATTTTCATTAAGCCAGTATAAATAGAACTAGCTTTTAAACTGACCTTTGGAACCAT 29660  
QY 2159 TTGTTTCAGTTAGTTTACCTGGGTATTTTGTCTGATTCACCTGCCAATTTGTCACTTT 2218  
Db 29659 TTGTTTCAGTTAGTTTACCTGGGTATTTTGTCTGATTCACCTGCCAATTTGTCACTTT 29600  
QY 2219 TAAGACTTTTTTTTTTCCATATAGGAAGCCATTTATAGTAGTAAACTTTTACAAATCC 2278  
Db 29599 TAAGAC - TTTT TTTT TTTTCCATATAGGAAGCCATTTATAGTAGTAAACTTTTACAAATCC 29541  
QY 2279 CATTTTCAAATTTACTTTTATGATCTTAAATTTTCAATTTTCTGCTAAATAACAGTGCCTCTA 2338  
Db 29540 CATTTTCAAATTTACTTTTATGATCTTAAATTTTCAATTTTCTGCTTAATAACAGTGCCTCTA 29481  
QY 2339 CCTTTTGACATCTGGCTCAATTTAAATAATTTAGCAATAGAAATGTAATTTGTAATAAAGTT 2398  
Db 29480 CCTTTTGACATCTGGCTCAATTTAAATAATTTAGCAATAGAAATGTAATTTGTAATAAAGTT 29421

QY 2399 TGTGTAATTAACCTCAAGGGTTTAAATTTTCTTACTAGCTTCTAAATGGATTAATAATCAAGT 2458  
Db 29420 TGTGTAATTAACCTCAAGGGTTTAAATTTTCTTACTAGCTTCTAAATGGATTAATAATCAAGT 29361  
QY 2459 GCTTCAAATGAATTAAGAGTCCAGTTTCGGAAGATAAATAATGTTTGTAGTACACCAT 2518  
Db 29360 GCTTCAAATGAATTAAGAGTCCAGTTTCGGAAGATAAATAATGTTTGTAGTACACCAT 29301  
QY 2519 AATTTTCAGATCAGTATATTTCTGAAGACTCTCTGTGTCTGGCTAAATAATTTGGCCATCTT 2578  
Db 29300 AATTTTCAGATCAGTATATTTCTGAAGACTCCCTGTGTCTGGCTAAATAATTTGGCCATCTT 29241  
QY 2579 TATTAATGAGCCCTTTAAGGAATAAACAACCTTAAACAAGCATCAGTATTTATAGCAAAA 2638  
Db 29240 TATTAATGAGCCCTTTAAGGAATAAACAACCTTAAACAAGCATCAGTATTTATAGCAAAA 29181  
QY 2639 AGAGACTCTGTTAGTGACATGCGCATTTTCGTCTCACTTAATAGTTGGCCCTAAATTAGTA 2698  
Db 29180 AGAGACTCTGTTAGTGACATGCGCATTTTCGTCTCACTTAATAGTTGGCCCTAAATTAGTA 29121  
QY 2699 CACAGGATATTTTGTGTTTTCATCTTCTTAAACATGCTATCTTTTCATTTAAATAATAG 2758  
Db 29120 CACAGGATATTTTGTGTTTTCATCTTCTTAAACATGCTATCTTTTCATTTAAATAATAG 29061  
QY 2759 TAAATAGTATGGCATTTGGGTCTTCAGAGTCGATATATAGTAGATCTCTTTAGTCTTTT 2818  
Db 29060 TAAATAGTATGGCATTTGGGTCTTCAGAGTCGATATATAGTAGATCTCTTTAGTCTTTT 29001  
QY 2819 TCCACCTTTTCATCCCAAGGGTGGTCAAGTGCAGCAGCAATTTATTTTCATTTGTTGG 2878  
Db 29000 TCCACCTTTTCATCCCAAGGGTGGTCAAGTGCAGCAGCAATTTATTTTCATTTGTTGG 28941  
QY 2879 CCCACGGTTAGTCCATAATCTAGAGCCATTTGGAACCTGCAGCCATGAGGTGTGTTATC 2938  
Db 28940 CCCACGGTTAGTCCATAATCTAGAGCCATTTGGAACCTGCAGCCATGAGGTGTGTTATC 28881  
QY 2939 CCACAGTGGATTTGACTCAGCCCTCTGTGGGTGACAGACTTCTAAGCAGGAAGATACAGTG 2998  
Db 28880 CCACAGTGGATTTGACTCAGCCCTCTGTGGGTGACAGACTTCTAAGCAGGAAGATACAGTG 28821  
QY 2999 AAGCACATGTTTACATTTTGGGAACCTTGTGTAGGATCATGCCCTGTAGCCAGGTTAA 3058  
Db 28820 AAGCACATGTTTACATTTTGGGAACCTTGTGTAGGATCATGCCCTGTAGCCAGGTTAA 28761  
QY 3059 AAACCTGGACTTTTGAAGTAAAGTAAAGCATATCGCTTATATCATTTCTTGTGAAT 3118  
Db 28760 AAACCTGGACTTTTGAAGTAAAGTAAAGCATATCGCTTATATCATTTCTTGTGAAT 28701  
QY 3119 TGATATGTTTTTCTTCCCTTAAAGATCAAAAGCAGAAAAACAAACACAGTCTACTC 3178  
Db 28700 TGATATGTTTTTCTTCCCTTAAAGATCAAAAGCAGAAAAACAAACACAGTCTACTC 28641  
QY 3179 CGATGTTATCTTCTGATTCATCTGATCCATCTTCTTCTGCAATATTTTGGATGGAGA 3238  
Db 28640 CGATGTTATCTTCTGATTCATCTGATCCATCTTCTTCTGCAATATTTTGGATGGAGA 28581  
QY 3239 ATTTGGAAGTTAAATGATTAAGAAACCTGATGAACTACCAACAAGTATTTTAAAGTACT 3298  
Db 28580 ATTTGGAAGTTAAATGATTAAGAAACCTGATGAACTACCAACAAGTATTTTAAAGTACT 28521  
QY 3299 AGAATATATACAGTAAATCCCACTTTTCATGCAATCTCTGGGAATATAGAGTATTCG 3358  
Db 28520 AGAATATATACAGTAAATCCCACTTTTCATGCAATCTCTGGGAATATAGAGTATTCG 28461  
QY 3359 AAATAAGTTCAGTTTGTAGAGGGTAAACAAGTAAAGTAAACCACTATCTTGGTTAAC 3418  
Db 28460 AAATAAGTTCAGTTTGTAGAGGGTAAACAAGTAAAGTAAACCACTATCTTGGTTAAC 28401  
QY 3419 ATGAAAAATAACAATTTGAGAATATATATATTTACTGAAATATATAGCTTTTCTCACA 3478  
Db 28400 ATGAAAAATAACAATTTGAGAATATATATATTTACTGAAATATATAGCTTTTCTCACA 28341

QY	3479	TTAGACAACCAACATACTCTTAAAGTGCTAAATTAATATATATTTTCTAAGGTCAGTT	3538	Db	27262	ACACACACACATACATAATCTTGCCCTGCATGAATTCAAATACATGAGGCACATCTT	27203
Db	28340	TTAGACAACCAACATACTCTTAAAGTGCTAAATTAATATATATTTTCTAAGGTCAGTT	28281	QY	4619	CAGGCGCACAGTGTAAAAATTTTGAGTCTTAATTTTCATGTGTACACCTCTTTGCTGCT	4678
QY	3539	GGGACATTAACCTAAGAAACATATCTTATTAAAGCATTGTAAACACCTTATTTTAGGACCC	3598	Db	27202	CAGGCGCACAGTGTAAAAATTTTGAGTCTTAATTTTCATGTGTACACCTCTTTGCTGCT	27143
Db	28280	GGGACATTAACCTAAGAAACATATCTTATTAAAGCATTGTAAACACCTTATTTTAGGACCC	28221	QY	4679	TCCCAACCCCGACACTTGAATAACACTTCAGAGTAAGAGGGAATTCAGCTAATTTGTTT	4738
QY	3599	TTTCGTTGGGATGGGGCAAGGTTGGAGGTTTGTGAAGATATATATCTCTTTAAA	3658	Db	27142	TCCCAACCCCGACACTTGAATAACACTTCAGAGTAAGAGGGAATTCAGCTAATTTGTTT	27083
Db	28220	TTTCGTTGGGATGGGGCAAGGTTGGAGGTTTGTGAAGATATATATCTCTTTAAA	28161	QY	4739	TAAATTTGACTGTAGTGGTCACTAAACCCCTTTTGAGGAATTTCTATTAAAGATGAGGC	4798
QY	3659	AAAAACAGAAAGAAATATTTCTGAGCACTCATTTAGCCCTATATGGAACCTCTTTCC	3718	Db	27082	TAAATTTGACTGTAGTGGTCACTAAACCCCTTTTGAGGAATTTCTATTAAAGATGAGGC	27023
Db	28160	AAAAACAGAAAGAAATATTTCTGAGCACTCATTTAGCCCTATATGGAACCTCTTTCC	28101	QY	4799	AGACTCGCTTATTTGAATTTGCAATGTTCTAACAGGATGTAAACACAGAAATGGCTTTT	4858
QY	3719	TTTTTGTAGGGCAGTTTCACTGCAAGATGCAATGTTTACCAGAAATTTCTAAATAACA	3778	Db	27022	AGACTCGCTTATTTGAATTTGCAATGTTCTAACAGGATGTAAACACAGAAATGGCTTTT	26963
Db	28100	TTTTTGTAGGGCAGTTTCACTGCAAGATGCAATGTTTACCAGAAATTTCTAAATAACA	28041	QY	4859	TTTTTCCCTAGAAAAAGATTTGTTTCTATGTCAACTAGATATGATTTAAAAATAAGTAT	4918
QY	3779	GTGCAGATTACTGAATATATACATTTTAAATATTTGGAGTAGTATATTTGTTGA	3838	Db	26962	TTTTTCCCTAGAAAAAGATTTGTTTCTATGTCAACTAGATATGATTTAAAAATAAGTAT	26903
Db	28040	GTGCAGATTACTGAATATATACATTTTAAATATTTGGAGTAGTATATTTGTTGA	27981	QY	4919	TGCCAATGCTGTTTCATTTCTAGTGGCCAGAACTATTATCTCTTGAAATTTCTGTGTAGT	4978
QY	3839	GAAATGTAATTTGAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG	3898	Db	26902	TGCCAATGCTGTTTCATTTCTAGTGGCCAGAACTATTATCTCTTGAAATTTCTGTGTAGT	26843
Db	27980	GAAATGTAATTTGAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG	27922	QY	4979	GCCTTAGCTTGGTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	5038
QY	3899	ACACACATGACACATACCGCACTTCATAGAAATCAAAAGTTGCTCTCTGAAGGAGCTTT	3958	Db	26842	GCCTTAGCTTGGTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	26784
Db	27921	ACACACATGACACATACCGCACTTCATAGAAATCAAAAGTTGCTCTCTGAAGGAGCTTT	27862	QY	5039	AGTAGTTTGAATTTGGGCTCAGACAGATATTGAACCTCATTTCACTGTTTCACTCCACA	5098
QY	3959	GGCTCTGATATTTTATGATGCTCTATATTTTAACTTTGGAGCAGTAGTTTTAT	4018	Db	26783	AGTAGTTTGAATTTGGGCTCAGACAGATATTGAACCTCATTTCACTGTTTCACTCCACA	26724
Db	27861	GGCTCTGATATTTTATGATGCTCTATATTTTAACTTTGGAGCAGTAGTTTTAT	27802	QY	5099	TGTATGTACAAGTTAGTTCACCAACACGGAAGT-TGAGTGGGAAGAGTCTTTGGCAGTG	5157
QY	4019	ACTTATGTTTAAATTTTATATGAAAAATACATTTTAAAAAAGTGTGTCCAAAG	4078	Db	26723	TGTATGTACAAGTTAGTTCACCAACACGGAAGTGTGAGTGGGAAGAGTCTTTGGCAGTG	26664
Db	27801	ACTTATGTTTAAATTTTATGAAAAATACATTTTAAAAAAGTGTGTCCAAAG	27742	QY	5158	TAGCAATGCTATCCATTCATGATATACAGTACCTTTATAGTTATCGATCACTGTTTAAAA	5217
QY	4079	GCATTAATAATPATATGTTAATGAAGAGTACATTTTAAATTTTCAAACTGCTCCTA	4138	Db	26663	TAGCAATGCTATCCATTCATGATATACAGTACCTTTATAGTTATCGATCACTGTTTAAAA	26604
Db	27741	GCATTAATAATPATATGTTAATGAAGAGTACATTTTAAATTTTCAAACTGCTCCTA	27682	QY	5218	CTTTTCAATTTTAAATCTTATACCAAGTTTCACTGTTTAAAAAATTCATTTGCTCGCTG	5277
QY	4139	GCCTTTGATAGGAGATATTTTCTGAAGTAGGCTTTGCTCTGCTCTCAATTAATGCTC	4198	Db	26603	CTTTTCAATTTTAAATCTTATACCAAGTTTCACTGTTTAAAAAATTCATTTGCTCGCTG	26544
Db	27681	GCCTTTGATAGGAGATATTTTCTGAAGTAGGCTTTGCTCTGCTCTCAATTAATGCTC	27622	QY	5278	ATTATGCATCACTCTGCTGCAACTTTTTTATTTCAATTTTAGTGTGTTTCTTTCAAGCTGTGT	5337
QY	4199	TTCTTTTAGTTTCTATGAACAGATTTGCTTACCTAAATCTTTAGTTGAATGATTAGTGT	4258	Db	26543	ATTATGCATCACTCTGCTGCAACTTTTTTATTTCAATTTTAGTGTGTTTCTTTCAAGCTGTGT	26484
Db	27621	TTCTTTTAGTTTCTATGAACAGATTTGCTTACCTAAATCTTTAGTTGAATGATTAGTGT	27562	QY	5338	ATTTTGGCTATTTGTTGCTTGTGCTTTATTTTCTTAGTCACTTTCTGGAATATAGTAT	5397
QY	4259	CAATATTTGCTTTAATCACCATATAAAGGAAAAAATTTGGTGACAGACAAATAGAAA	4318	Db	26483	ATTTTGGCTATTTGTTGCTTGTGCTTTATTTTCTTAGTCACTTTCTGGAATATAGTAT	26424
Db	27561	CAATATTTGCTTTAATCACCATATAAAGGAAAAAATTTGGTGACAGACAAATAGAAA	27502	QY	5398	ATAATGTTTAAATTTGGACAGTAGCGGTTTTTAAAAACCATATACACTGCTGAAACATGAG	5457
QY	4319	ACCTATTTTAAATAGAAATCAAAATAGCAAGTGTGGAGCACTATTTATTTCTGTTTA	4378	Db	26423	ATAATGTTTAAATTTGGACAGTAGCGGTTTTTAAAAACCATATACACTGCTGAAACATGAG	26364
Db	27501	ACCTATTTTAAATAGAAATCAAAATAGCAAGTGTGGAGCACTATTTATTTCTGTTTA	27442	QY	5458	CCAGAGCCGATTTGCTTTTAAAGCTAATAATGTTTAAAGAGTACATATTTTCAGGAT	5517
QY	4379	AAATGTACTTAAAGATTCATCAATTTAGTGAACTGAGACATTTGGCCTTAGTGCTGTAT	4438	Db	26363	CCAGAGCCGATTTGCTTTTAAAGCTAATAATGTTTAAAGAGTACATATTTTCAGGAT	26304
Db	27441	AAATGTACTTAAAGATTCATCAATTTAGTGAACTGAGACATTTGGCCTTAGTGCTGTAT	27382	QY	5518	CGTTCATCTAGTCAGCAATACACATATATAGCCCAATATTTTTTAAAAAATAAGAGCTT	5577
QY	4439	TCATCTGCTAATTTAAAAAGGAGTACAGGATTTTATTAAGTAAAGCAATTTTGGAAATGG	4498	Db	26303	CGTTCATCTAGTCAGCAATACACATATATAGCCCAATATTTTTTAAAAAATAAGAGCTT	26244
Db	27381	TCATCTGCTAATTTAAAAAGGAGTACAGGATTTTATTAAGTAAAGCAATTTTGGAAATGG	27322	QY	5578	GGTCAACCTCTATACATACATATTAAGAGATATAGCCTTTTCAAAATGAATCTAAACCT	5637
QY	4499	GGATAGCCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	4558	Db	26243	GGTCAACCTCTATACATACATATTAAGAGATATAGCCTTTTCAAAATGAATCTAAACCT	26184
Db	27321	GGATAGCCCATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	27263	QY	5638	TTACAGAAACTTTTATAGGTTATGCTTTTATTTTAAAGACTTATTTATTTCAAGTGC	5697
QY	4559	ACACACATACATCTTAAATCTTGCCCTGCATGAAATTTCAATATACATGAGGCACATCTT	4618				

Db 26183 TTACAGAACTTCTTATAGGTTAAGCCTTTATTTTAAAGCTTATTATATCAAGTGC 26124  
QY 5698 CATTAGATGATATATATAGTGGCTTTGATATATATGCTTTCTGACAAAATGTTAGA 5757  
Db 26123 CATTAGATGATATATATAGTGGCTTTGATATATATGCTTTCTGACAAAATGTTAGA 26064  
QY 5758 TGTATTTTAAACAGGTACATTTTAAACAGTGTCTTATATCAATTTGCTATATTGACACAG 5817  
Db 26063 TGTATTTTAAACAGGTACATTTTAAACAGTGTCTTATATCAATTTGCTATATTGACACAG 26004  
QY 5818 AATCAGTGTGTCCTTTTCAATAGGTTTCAATAGTGTCTTATATCAATTTGCTATATTGACACAG 5877  
Db 26003 AATCAGTGTGTCCTTTTCAATAGGTTTCAATAGTGTCTTATATCAATTTGCTATATTGACACAG 25944  
QY 5878 GTCTCAAGCACACTGCTCTCCAGTACGTAGTTTAAATAATACCAAGTTTCAAGTTG 5937  
Db 25943 GTCTCAAGCACACTGCTCTCCAGTACGTAGTTTAAATAATACCAAGTTTCAAGTTG 25884  
QY 5938 CTTCTAGCCTACTGAGATCCATGTGACATTTGGAGGAGATCTTTTAAATGTTTAGTATTGCG 5997  
Db 25883 CTTCTAGCCTACTGAGATCCATGTGACATTTGGAGGAGATCTTTTAAATGTTTAGTATTGCG 25824  
QY 5998 TCATTAGCAATGGCTGGCTGTAGTTCTGGTAAATGTGTCCTAAAGTTGCAATTTGCTTGG 6057  
Db 25823 TCATTAGCAATGGCTGGCTGTAGTTCTGGTAAATGTGTCCTAAAGTTGCAATTTGCTTGG 25764  
QY 6058 TTTTCTCACACTGTGTCAGCAGCCATGCTCAACACAGATAGTCTGTTGTGATCACA 6117  
Db 25763 TTTTCTCACACTGTGTCAGCAGCCATGCTCAACACAGATAGTCTGTTGTGATCACA 25704  
QY 6118 TAGATCTACATAAGTTGTGAGTTTGTCTTAAATCCATAGGGAGCTCTTTTGGGATC 6177  
Db 25703 TAGATCTACATAAGTTGTGAGTTTGTCTTAAATCCATAGGGAGCTCTTTTGGGATC 25644  
QY 6178 ATAGAAAAGAGATCATGCAACAGCATTTGGTGAAGGCACTCAGATTGCACTTAGGGC 6237  
Db 25643 ATAGAAAAGAGATCATGCAACAGCATTTGGTGAAGGCACTCAGATTGCACTTAGGGC 25584  
QY 6238 CTTTCTATGATGTGTCAACCTCTGAGGATGGAGGAGTCTTTTGTATCTAG 6297  
Db 25583 CTTTCTATGATGTGTCAACCTCTGAGGATGGAGGAGTCTTTTGTATCTAG 25524  
QY 6298 CCTAGAAATGACACAGAACTATTGCTTAATGTATATAAAACACTTCATTATATAAGCTTCAGT 6357  
Db 25523 CCTAGAAATGACACAGAACTATTGCTTAATGTATATAAAACACTTCATTATATAAGCTTCAGT 25464  
QY 6358 GGTACAGATGAACACAGAAATGTTTCTCTCAGAAAACACTCCTTCAATATTATATT 6417  
Db 25463 GGTACAGATGAACACAGAAATGTTTCTCTCAGAAAACACTCCTTCAATATTATATT 25404  
QY 6418 GGATCATGCTGTAATGTAACCTTGGGCTACAACTCTTCATGCTGCTACAACTTCTCTGT 6477  
Db 25403 GGATCATGCTGTAATGTAACCTTGGGCTACAACTCTTCATGCTGCTACAACTTCTCTGT 25344  
QY 6478 CTCAFTCAGTGTATTTTATCCATAGAAAAGCACTACATTAGGTGTAAGAGTGAC 6537  
Db 25343 CTCAFTCAGTGTATTTTATCCATAGAAAAGCACTACATTAGGTGTAAGAGTGAC 25284  
QY 6538 AATATATTTTATACGTGACTTAATTTGTCAATTAACAACTTTTACACCAACCAATGT 6597  
Db 25283 AATATATTTTATACGTGACTTAATTTGTCAATTAACAACTTTTACACCAACCAATGT 25224  
QY 6598 ATTCATGTCATCTGCAAAAGGAGATCTCGGACATGCAAAAGTTACAGAACCAACCCAG 6657  
Db 25223 ATTCATGTCATCTGCAAAAGGAGATCTCGGACATGCAAAAGTTACAGAACCAACCCAG 25164  
QY 6658 CTTTGTGCCAAGGTGACTGTAACTCAGATGGAAGTGGGCTTTTATAATAGGTTGG 6717  
Db 25163 CTTTGTGCCAAGGTGACTGTAACTCAGATGGAAGTGGGCTTTTATAATAGGTTGG 25104  
QY 6718 AGTGAAGAACATGCTGTATGTTACTAAACAGCCCTTTGAATTTTAAACAAATCTGGGAATCC 6777  
Db 25103 AGTGAAGAACATGCTGTATGTTACTAAACAGCCCTTTGAATTTTAAACAAATCTGGGAATCC 25044

QY 6778 ATTAGAAAACGATTTGCTCATCACTACCTGAACATTAAGCTGCACTGCTGAAATTTGATTTT 6837  
Db 25043 ATTAGAAAACGATTTGCTCATCACTACCTGAACATTAAGCTGCACTGCTGAAATTTGATTTT 24984  
QY 6838 GCTAATGAAAAGTGTTCGACTAGTACTCTAAATGTTTCTAAATGATTAAGTTTGGT 6897  
Db 24983 GCTAATGAAAAGTGTTCGACTAGTACTCTAAATGTTTCTAAATGATTAAGTTTGGT 24924  
QY 6898 CAAAATAGAAAAGAAAATCTGCAATTCAGGCCGAAATTTTGTATATTTTATTTGCAAT 6957  
Db 24923 CAAAATAGAAAAGAAAATCTGCAATTCAGGCCGAAATTTTGTATATTTTATTTGCAAT 24864  
QY 6958 TAAAATGCTATCTGTAATATTTGGGAAATCAAGTGGCTTATCATGTATATCGTGTACT 7017  
Db 24863 TAAAATGCTATCTGTAATATTTGGGAAATCAAGTGGCTTATCATGTATATCGTGTACT 24804  
QY 7018 AAAATGCTATCTGTAATATTTGGGAAATCAAGTGGCTTATCATGTATATCGTGTACT 7077  
Db 24803 AAAATGCTATCTGTAATATTTGGGAAATCAAGTGGCTTATCATGTATATCGTGTACT 24744  
QY 7078 TGTGTGTATTAAGCTCTGTAATTAAGCAATCACATTATGAAGCTGCACTGATACAT 7137  
Db 24743 TGTGTGTATTAAGCTCTGTAATTAAGCAATCACATTATGAAGCTGCACTGATACAT 24684  
QY 7138 TTTAAACATTCATCTCCAAAGAGCAGACTATTTATGTGCCATATACCAAGTTTAAATA 7197  
Db 24683 TTTAAACATTCATCTCCAAAGAGCAGACTATTTATGTGCCATATACCAAGTTTAAATA 24624  
QY 7198 TTAATTTGCTGCTAATTAATTAATAGTACTGAGCTTCTTGTGGCTACAGTGTATGTT 7257  
Db 24623 TTAATTTGCTGCTAATTAATTAATAGTACTGAGCTTCTTGTGGCTACAGTGTATGTT 24564  
QY 7258 TGCTTAAGATAAGATATGTAATTCACAAAATATATGAATAAAAT 7305  
Db 24563 TGCTTAAGATAAGATATGTAATTCACAAAATATATGAATAAAAT 24516

RESULT 4  
AC097274/c

LOCUS AC097274 182102 bp DNA linear HTG 18-JAN-2002  
DEFINITION Mus musculus clone RP23-59B17, WORKING DRAFT SEQUENCE, 8 unordered pieces  
ACCESSION AC097274  
VERSION AC097274.5 GI:18201773  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 182102)  
AUTHORS McCombie, W.R., Baker, J.P., Balija, V., Dedhia, N.N., de la Bastide, M., Kuit, K., King, L., Kirchoff, K.A., Miller, B., Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R., Santos, L., Shah, R.S., Spiegel, L.A., Palmer, L., Yang, C. and Zutavern, T.  
TITLE Mouse Genomic Sequence  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 182102)  
AUTHORS McCombie, W.R.  
TITLE Direct Submission  
JOURNAL Submitted (13-OCT-2001) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA  
COMMENT On Jan 18, 2002 this sequence version replaced gi:16973738.  
----- Genome Center  
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory  
Center code: CSHL  
Web site: <http://www.cshl.org/genseq>  
Contact: [mccombie@cshl.org](mailto:mccombie@cshl.org)  
----- Project Information  
Center project name: RP23-59B17  
Center clone name: RP23-59B17

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 48370: contig of 48370 bp in length
* 48371 49082: gap of unknown length
* 49083 95033: contig of 45951 bp in length
* 95034 95745: gap of unknown length
* 95746 129089: contig of 3324 bp in length
* 129090 129781: gap of unknown length
* 129782 150082: contig of 20301 bp in length
* 150083 150793: gap of unknown length
* 150794 162439: contig of 11646 bp in length
* 162440 163150: gap of unknown length
* 163151 172039: contig of 8949 bp in length
* 172100 172810: gap of unknown length
* 172811 178597: contig of 5786 bp in length
* 178598 182102: gap of unknown length
* 182103 182102: contig of 2795 bp in length.
  Location/Qualifiers
    1. 182102
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10090"
      /clone="RP23-59B17"
  BASE COUNT 50661 a 36394 c 36240 g 53815 t 4992 others
  ORIGIN

```

```

Query Match      38.6%; Score 2821.6; DB 2; Length 182102;
Best Local Similarity 73.1%; Pred. No. 0;
Matches 4933; Conservative 0; Mismatches 1294; Indels 525; Gaps 81;

QY 719 AAAGCAGATGGCTACTCATCTCTCCGAGAAACCCACAGAGTGAATTTGAGAGAA 778
DB 60773 AAAGCAGATGGCTACTCATCTCTCCGAGAAACCCACAGAGTGAATTTGAGAGAA 60714

QY 779 TGTTTACCAGGAAAGATCATCTGAAGAAATCACCTCCATACACAGAGCCCTTAACAAAGAGA 838
DB 60713 TGTTTACCAGGAAAGATCATCTGAAGAAATCACCTCCATACACAGAGCCCTTAACAAAGAGA 60654

QY 839 CGTTTAAGTCGGAAGATGTGGCAAGAACTACAAATACCAAGCTTGAATTTAAACGTCAC 898
DB 60653 CCTTTAAATCGGAAGAGTGTGGCAAGAGCTACAAATACCAAGCTTGGTTTAAAGCGACACT 60594

QY 899 TGCCCTTGCATGCGCAAGAGTGTGACCTCACCTGTAAGGTATGTTTGCAGAACTTTTG 958
DB 60593 TGCCCTTGCATGCGCAAGAGTGTGACCTCACCTGTAAGGTATGTTTGCAGAACTTTTG 60534

QY 959 AAAGCAGGAGTGTCTTCTGGAGCAGCTTAAATCTCATGAGGCAAGTCTGTGGTGGG 1018
DB 60533 AGAGCAGAGTGTCTTCTGGAGCAGCTTAAATCTCATGAGGCAAGTCTGTGGTGGG 60474

QY 1019 TTAAGAGAAAAGCAGCAGTGGCAACATTTGTGATGTCGCGGTCTACACCCGAAAGATG 1078
DB 60473 TGAAGGAGAAAAGCAGCAGTGGCAACATTTGTGATGTCGCGGTCTACACCCGAAAGATG 60414

QY 1079 TCCGGAGACATGTTGTCACACTGGAAGAAAGAGTCTTCTCTGTGAGTATTTGTCAC 1138
DB 60413 TCCGGAGACATGTTGTCACACTGGAAGAAAGAGTCTTCTCTGTGAGTATTTGTCAC 60354

QY 1139 AGAGATTGGCGGAAGGATCACCTGATCTGCATATGAGAGAGATCAACAATCAAGAGC 1198
DB 60353 AGAGATTGGCGGAAGGATCACCTGATCTGCATATGAGAGAGATCAACAATCAAGAGC 60294

QY 1199 TTCTGAAGGTCAAAACAGAACAGTGTGATTTCTTGACCCATTTTACCTGCAATGTGCTG 1258
DB 60293 TTCTGAAGGTCAAAACAGAACAGTGTGATTTCTTGACCCATTTTACCTGCAATGTGCTG 60234

```

```

QY 1259 TGCCTATAAAGACGAGCTCTCTTCGGGTGATGCTCTTACCTTCCAGTGAACCTGTATCA 1318
DB 60233 TGCCTATAAAGATGAACCTCTTCGGGTGATGCTCTTACCTTCCAGTGAACCTGTATCA 60174

QY 1319 AGCCATTCAAAACACTTTTGGAGTTAAACCTCTACACACTCCATTTTCACTCAATGAGA 1378
DB 60173 AGCCATTCAAAACACTTTTGGAGTTAAACCTCTATACACTCCATTTTCACTCAATGAGA 60114

QY 1379 GCTCGGATCTGCCACCAAAATGATCAACAACTTTTACCTTTGGGAATGACATGCCCAATAG 1438
DB 60113 GCTCGGATCTGCCACCAAAATGATCAACAACTTTTACCTTTGGGAATGACATGCCCAATAG 60054

QY 1439 ATATGACACTGTTTATCTCCCTCTCACCACTTTTCTTTTCAAAATCCGTTTCACTTACCT 1498
DB 60053 ATATGACACTGTTTATCTCCCTCTCACCACTTTTCTTTTCAAAATCCGTTTCACTTACCT 59994

QY 1499 CATATGCAATTTCTATTTCTGAAAGAAAGACAGCATTAAGGGGGGAAATGAGAGTTACC 1558
DB 59993 CATATGCAATTTCTATTTCTGAAAGAAAGACAGCATTAAGGGGGGAAATTTGAGAGTTACC 59934

QY 1559 TGATGGAGTTACAAGGTGGGTGCGCTCTTATCCCAAGATTCTCAAGCATCTCTCATCAT 1618
DB 59933 TGATGGAGTTACAAGGTGGGTGCGCTCTTATCCCAAGATTCTCAAGCATCTCTCATCAT 1678

QY 1619 CTAAGCTAGGTTGGATCTCTCAGATTGGGTCCCTAGATGATGGTGCAGGAGACTCTCCC 1678
DB 59876 CTAAGCTAGGTTAGAGCTCTCAGATTGGGTGCGCATGATGGTGGGAGACTCTCCC 59817

QY 1679 TATCCAAAAGCTTATCTCTCAGTGCAGTGCCTTAACACACACAGCATTTGGATTTTCTC 1738
DB 59816 TGTCAAAGAGCTCTATCTCTATCAGTGACCCCTCAGCACACACAGCATTTGGATTTTCTC 59757

QY 1739 AGTTGTTTAAATTTTATACCTTTTAAATGGTCTCTCTCTATATCTCTATCAGTGGGAGCC 1798
DB 59756 AGTTGTTTAAATTTTATACCTTTTAAATGGTCTCTCTCTATATCTCTATCAGTGGGAGCC 59697

QY 1799 TTGGAATGAGCTATTTCCAGGAGAAAGACATTTCTTCTTTTCCAGCTCCCCACACAAA 1858
DB 59696 TTGGAATGAGCTATTTCCAGGAGAAAGACATTTCTTCTTTTCCAGCTCCCCACACAAA 59637

QY 1859 CACAGATCTTCHAGATCTCTGAAACACTATAGGCTTGGGTCTCTGCACTCATCTCTCAG 1918
DB 59636 CACAGATCTTCHAGATCTCTGAAACACTATAGGCTTGGGTCTCTGCACTCATCTCTCAG 59577

QY 1919 CAGCTTTTACCAGCAGTTTAAAGCACAAGTACCACCTCCACAGTTTCCATCAAGCTTTTC 1978
DB 59576 CAGCTTTTACCAGCAGTTTAAAGCACAAGTACCACCTCCACAGTTTCCATCAAGCTTTTC 59517

QY 1979 AGTAGGATTTCTGGGACATGGAATTTATACAGAAATGATGTGTAGTGTGCTTCCCTAGATG 2038
DB 59516 AGTAGGATTTCTGGGACATGGAATTTATACAGAAATGATGTGTAGTGTGCTTCCCTAGATG 59459

QY 2039 CAAATTTTATTTAGTGGCTTAC - TTTAAACAGATATAAAATTTCTGCTTTTGTATAATA 2097
DB 59458 CAAATTTTATTTAGTGGCTTAC - TTTAAACAGATATAAAATTTCTGCTTTTGTATAATA 59399

QY 2098 CAAATTTTCAATTAAGCCAGTATAAAATAGAAATAGCTTTTAAACTGAGCTTTTGGAAACCA 2157
DB 59399 CAAATTTTCAATTAAGCCAGTATAAAATAGAAATAGCTTTTAAACTGAGCTTTTGGAAACCA 59339

QY 2157 TTTGTGTTCAAGTTAAGTTTACCTGGGTATTTTGTCTGATTCATCTGCAATTTGTCACTT 2217
DB 59338 TTTGTGTTCAAGTTAAGTTTACCTGGGTATTTTGTCTGATTCATCTGCAATTTGTCACTT 59285

QY 2218 TTAAGACTTTTTTTTTTCCATATAGAAAGCCATTTATAGTAGTAACTTTTACAAATC 2277
DB 59284 TTAAGACTTTTTTTTTTCCATATAGAAAGCCATTTATAGTAGTAACTTTTACAAATC 59230

QY 2278 CCAATTTTCAATTTTACATTTTAAATTTTCAATTTTGTCTTAAATACAGTGGCTCT 2337
DB 59229 CCAATTTTCAATTTTACATTTTAAATTTTCAATTTTGTCTTAAATACAGTGGCTCT 59172

QY 2338 ACCTTTTGACATCTGGCTCATTTAAATAATTTAGCAATAGATGTAAATTTGTATATAAAG - 2396

```







```
Db 55118 AGTGATTTCATGTCGCTTGCACAAAGGAGATCTTGGATGTGCAATGTTACCGAACA 55059
QY 6653 CCAGCTTTTGTCCACAAAGTGACTGTAATCTCAGATGGAAGTGGCTTTATAATAGG 6712
Db 55058 CCAGCTTTTGTCCACAAAGTGACTGTAATCTCAGATGGAAGTGGCTTTGA-CATAGG 55000
QY 6713 TGTGGAGTGAAGAACATGCTGTGTATGTTACTAAACAGCCCTTTGAATTTAAACAAAACCTGG 6772
Db 54999 TGTGGAGTGAAGAACATGCTGTGTATGTTACTAAACAGCCCTTTGAATTTAAACAAAACCTGG 54940
QY 6773 AATCCATTAG-GAAACGGATTGATCATACCTGAAACATAGCTG----- 6815
Db 54939 AATCCATTAGAGAAATGGATTGATCATGCTGAAATAGCTGACTGCTGTTTTTTT 54880
QY 6816 -----GACTGCTCAAAATGTTGATTTAGTCTAATGAAAGTGT 6853
Db 54879 TTGTTGTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 54820
QY 6854 TTGGACTAGTACTCTAAATGTTCTAATGATAAAGTTTGTAGTCAAAATAGAAAAGAA 6913
Db 54819 TTGGACTAGTACTCTAAATGTTCTAATGATAAAGTTTGTAGTCAAAATAGAAATG-AA 54761
QY 6914 AATCTGCAATTCAGCCGCAATTTTGTATATTTTATTGCAATTTAAATTTGCTATTCTG 6973
Db 54760 AATCTGCAATTCAGACTCAATTTTGTATATTTTATTGCAATTTAAATTTGCTACTG 54701
QY 6974 TAATATTGGAATCAAGTGGCTTATCATGTATATGCTACTTAAATGTTATTCAAA 7033
Db 54700 TGATATTGGAAGCAAGTGGCTTATCATGTATATGCTACTTAAATGTTATTCAAA 54641
QY 7034 CTACTGTTGTTATTTGTTATATAATATAGACAAAGATCATATTTTGTGTGTATAGCT 7093
Db 54640 CTACTGTTGTTATTTGTTATATAATATAGACAAAG-----TTATCTGTGTGTATACCT 54588
QY 7094 CTGTAATAATAGCAATCAATATGAAAGTGCAGTGATATCAATTTAAACATTCATC 7153
Db 54587 CTGTACAGTTGC-TTCACGCCATGAAACGTCATATCCATAGCTGAGAAACATACAC 54529
QY 7154 CAAAGACGAGACTATTTATGTCATATACCATACCATGTTAAATATTAATTTGCTGCTAAT 7213
Db 54528 CAAAGACGAGGATGCTTTATGTCACACACCAAGCTGCACAGTCATCTGCTG----- 54473
QY 7214 TAAATAATAGTACTGAGCTCTTGTGGCTACAGTGTATGTTGCTGTAGAAATAGA 7273
Db 54472 TGCCGCTTACCACACAGCTGCTTATGACCTGCAGTGGGATTTGTTGT-AGGAAAAA 54414
QY 7274 TATGTAATTCACAAAATATATGATAAAAT 7305
Db 54413 TGGATGAATTCGACAGATACTTGATTTGAAT 54382
```

```
RESULT 5
AL807387/c
LOCUS Mouse DNA sequence from clone RP23-24J10 on chromosome 4, complete
DEFINITION
ACCESSION AL807387
VERSION AL807387.10 GI:30350042
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 241148)
Harrison, E.
Direct Submission
Submitted (03-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On May 3, 2003 this sequence version replaced gi:26788130.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
```

from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-24J10 is from the RPCI-23 Mouse BAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

#### FEATURES

##### source

1..241148  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-24J10"  
/clone\_lib="RPCI-23"  
BASE COUNT 71321 a 48483 c 49088 g 72256 t  
ORIGIN

##### Query Match

Best Local Similarity 38.6%; Score 2821.6; DB 10; Length 241148;  
Matches 4933; Conservative 0; Mismatches 1294; Indels 525; Gaps 81;

QY	719	AAAGGCACATGGCTACTCATCTCTCTGAGAAACCCACAAAGTGAATTTATTTCTGAGAAA	778
Db	75206	AAAGGCACATGGCTACTCATCTCTCTGAGAAACCCACAAAGTGAATTTATTTCTGAGAAA <td>75147</td>	75147
QY	779	TGTTTACCGGAAAGATCATCTGAAGAATCACTCCATACACACGACCTTAAACAAAGAGA	838
Db	75146	TGTTTACCGGAAAGATCATCTGAAGAATCACTCCATACACATGACCCCAACAAAGAGA	75087
QY	839	CGTTTAAAGTGCAGAAATGCGCAGAACTACATACCAAGCTTGAATTTAAACGCTACT	898
Db	75086	CCTTTAAAGTGCAGAGTGTGGCAGAGCTACATACCAAGCTTGGTTTAAAGCAGCTACT	75027
QY	899	TGGCCTTGCATGCCCAACAAAGTGGTGAGCTCACTCACTGTAAGGTATGTTTGCACCTTTTG	958
Db	75026	TGGCCTTGCATGCTGCCACCAAGTGGTGAGCTCACTCACTGTAAGGTATGTTTGCACCTTTTG	74967
QY	959	AAAGCAGGGAGTGCTTCTGAGGACCTTAATCTCATGAGGCAAGTCTGTTCTGTTGGGG	1018
Db	74966	AGAGCACAGGTGTGCTCTCTAGAGCACCCTGAAATCTCACGCGGGCAAGTCATCTGGGGTG	74907
QY	1019	TTAAAGAAAAAAGCAGGTGCGAACAATTTGTGATCGCCGGTTCTTACACCCGAAAGATG	1078
Db	74906	TGAAGGAAAAAAGCAGCAATGTGAGCACTCCGCAAGCTTCTACACCCGGAAGATG	74847
QY	1079	TCCGAGACACATGGTGTGTGACACTGGAAGAAAGGACTTCTCTGTGATTTGTGCAC	1138
Db	74846	TCCGAGACACATGGTGTGTGACACTGGAAGAAAGGACTTCTCTGTGATTTGTGCAC	74787

QY	1139	AGAGATTTGGCGAAGGATCACCTGACATGACATATGAAGAGTCAATCAAGAGC	1198	QY	2218	TTAAGACTTTTTTTTCCATATAGGAAGCCATTTATTAGTAGTAAACATTTTACAAATC	2277
Db	74786	AGAGATTTGGAGAAAGGATCACCTCACGACATGAAGAGAGTCAATCAGGAGC	74727	Db	73717	CTCAGAGTTAAATTT---CATATAGGAGCGCAATTTATTATCACTAAA-TTTTACAAGTC	73663
QY	1199	TTCTGAAGTCAAAAAGAGAGAGTGGATTTCTGACCCATTTTACCTCAATGTCTG	1258	QY	2278	CCATTTTCAAAATCTTTTAGATCTTAAATTTTCAATTTTCTTAATAAACAGTGTCTCT	2337
Db	74726	TTCTGAAGTCAAAAAGAGAGAGTGGATTTCTGACCCATTTTACCTCAATGTCTG	74667	Db	73662	CTATGTTCAAATTTACTGTGTAGATCTTAAATTTTCAATTTTCTTAATAGCAATGGC--	73605
QY	1259	TGCTATAAAGACGAGCTCTTCCGGTGAATGCTTACCTTCCAGTGAATGTATCAA	1318	QY	2338	ACCTTTTGACATCTGGCTCAATTAATAAATTTAGCAATGAATGTATAAATAAG-2396	
Db	74666	TGCTATAAAGATGAATCTCTGCCGTGATGCTTACCTTCCAGTGAATGTATCAA	74607	Db	73604	----ATTGACCTTTTTTTTTTGTCAATATTCAGTACAGTGAATCTTATGAAGTGT	73549
QY	1319	AGCCATTTCAACAACCTTTGAGTTAACTTACCAACACTCCATTTCACTGTCATCAG	1378	QY	2397	---TTTGTGAATACTCAAGGTTTAA-ATTTCCTTACAGCTTCTTAAATGGATTAATAT	2453
Db	74606	AGCCATTTCAACAACCTTTGAGTTAACTTACCAACACTCCATTTCACTGTCATCAG	74547	Db	73548	GTGTGTGAATAACTCAAGTGTTCACATTTTTTTTACTAGCCACTAAATGGATTCATA	73489
QY	1379	GCTCGGATCTGCCACCAATGATCAACACTTTTACCTTTGGGAATGACATGCCCAATAG	1438	QY	2454	CAA--GTGCTTCAATGAATTAAGAGTCCAGTTTCGGAAGATAAATAATTTTGTAGAT	2511
Db	74546	GCTCTGGTCTGCTCAACCAATGATCAACACTTTTACCTTTGGGAATGACATGCCCAATAG	74487	Db	73488	TAAGTGTGTTTAAATGAGTTAAGATCCAGTTTGGGAGAGTATAGTGTTCATAAT	73429
QY	1439	ATATGACACTGTATCCCTCTCACACCTTCTTCAAAATATCCGTTTCACTTACCT	1498	QY	2512	ACACCAATAATTCAGATCAGTATATTTCTGAAGACTCTCTGTGTCTGGCTTAAATATTTG	2571
Db	74486	ATATGATGTGTCTACCCCTCTCATCTCTGCTTTCAAAATCCCATTCAGTTCACCT	74427	Db	73428	ATACCATCATTTTCAATCAATATTTTGGAGACT-GCTATTCTCTGGCTTAAATATTTT	73370
QY	1499	CATATCAATTTCTATTCTTCTGAAAGAGACAGCCATTTAAAGGGGAAATTTGAGATTACC	1558	QY	2572	CCATCTTTTATTATGAGCTTTTAAAGAAACAAACCCCTAAACACAAAGCATCAGTATTTAT	2631
Db	74426	CATACGCAATCTCTATTCTTCTGAAAGAGACAGCCATTTAAAGGGGAAATTTGAGATTACC	74367	Db	73369	-----AAAGTTTCAGTATCTAT	73354
QY	1559	TGATGAGTTTACAGGTGGCTGCCCTCTTTCATCCCAAGATTTCAAGCATCTGTCATCAT	1618	QY	2632	AGCAAAAAGAGACTCTTTAGTGACATGGCATTTCTGTCTCACTTAATAGTTGGCCCTAA	2691
Db	74366	TGATGAGCTTACAGGTGGTGCACCATCTTTCATCCCGAGGATTTCCAGCATCG---TCAT	74310	Db	73353	---CTAAAGAAACCTTAGAGAGGTAGCAGGTGTGTGTGTAGT-----	73312
QY	1619	CTAAGCTAGGTTGGATCTCTCAGATTGGTCCCTAGATGATGTGTCAGGAGACTCTCC	1678	QY	2692	ATTAGTACACAGGATATTTTGTGTCTTTCATCTTAAACATGCTATCTTTTCAATTA	2751
Db	74309	CTAAGCTAGGTTAGAGCTCTCAGATGGTTCGCGAGATGATGTGTGCGGAGCTCTCC	74250	Db	73311	ACTAGCCCTCAGCTGGTTTGGCGCATTTCAATCATCTTCCATA-----	73267
QY	1679	TATCCAAAAGCTCTATCTCCATCAGTGACCCCTTAAACACACAGCATTTGGATTTTCTC	1738	QY	2752	ATAATAGTAATAGTATGSCATTTGGGCTTTCAGAGTCGATATATAGGTAGATCTCTTT	2811
Db	74249	TGTCAAAGAGCTCTATCTCTATCAGTGACCCCTCAGCACACACAGCATTTGGATTTCTC	74190	Db	73266	-----AACAGTGTGGCATCGAAGTCATCAGGATCAGTGTAGAAAGTTTGTAT	73215
QY	1739	AGTTGTTTAAATTCATACCTTTAAATGGTCTCCCTATATCTCTATCAGTGGGAGCC	1798	QY	2812	AGTCTTTTCCACCTTTTCAATCCAGGGTGGGTCAAGTCAGCCAGCAGCAATTTATTTTCA	2871
Db	74189	AGTTGTTTCAATTTATACCAATTTGAATGGTCCCTTATAACCCACTTTTCAAGTGGAGCC	74130	Db	73214	ATGCCCTGTACCTTTTCAATACCCCA--GGATGGCTTCAGTATACAGGTTCTCATTTTGC	73157
QY	1799	TTGGAATGAGCTATTTCCAGGAGAGACATTTCTCTGTCTTCCAGCTCCACACACAAA	1858	QY	2872	TTGTGCGCCCAAGTTTGTAGTCCATAATC-TAGAGCCATTTTGGAACCTGCAGCCATGAGGTG	2930
Db	74129	TTGGATGAGCTATTTCCAGGAGAGACATTTCTCTGTCTCTCAGCTGCCACACAAA	74070	Db	73156	ATGGTGACACAGTATTTCCAGAGCTTAGAGCCATTTGGGACATAGACAGTAATA	73097
QY	1859	CACAGGATCTTCAAGGATCTTCAAAACACTATAGGGCTTGGTCTCTGCACTCACTGTGAG	1918	QY	2931	TGTTTATCCACAGTGGATGACTCAGCC---TCTGTGGGTGACAGACTTCTTAAGCAGGA	2987
Db	74069	CACAGGATCTTCAAGGATCTTCAAAACACTATAGGGCTTGGTCTCTGCACTCACTGTGAG	74010	Db	73096	TGTGAAACCTTACATGGCTGGGTGCTCAATCTGTGGGTGATAGACTCTCTCAACAGGA	73037
QY	1919	CAGCTTTTCAAGCAGTTTAAAGCACAGTACACCCCTCCACCGTTTCCATCAAGCTTTTC	1978	QY	2988	AGATAGCTGGAAGCACATGTTTACATTTTGGGAACCTTGTGTAGGATCATGGCCCTGTGA	3047
Db	74009	CAGCTTTCAAGCAGCTCAGCTCAAGCACTACTTTGCCCGTTTCCACAGGCGTTTC	73950	Db	73036	AGATGGATGTAAACA---GCCAAACCCAGAACCTTGTCCAGAGATT---CCACTGTT	72984
QY	1979	AGTAGGATCTGGGACATGATTCATTTACAGAAATGATGTAGCTGTGCCCTAGATGA	2038	QY	3048	GCCAGGTTTAAACCTGGAATTTTGAAGTAAGTAAAGTAAAGACATGAGCTTATATCATTT	3107
Db	73949	AGTAGGATCTGGGCGGAGGACTCA--CCGAGGTGCTTGTGTAGCCCTGCTGCAATGA	73892	Db	72983	GCCAGACTGAAACCTGAACTCTTGGTGCAGAGTAAAGCCTA---TCACATCACTT	72928
QY	2039	CCATTTTATTTTAGTCCCTTAC-TTTTAAACAGTATAAATAATTTCTGCTTTTGTATAATA	2097	QY	3108	CTTGCTGAATTTGATATGTTTTTCTTCCCTTTAAGATCAAAAGCAGAAACAAACAA	3167
Db	73891	CCATTTATTTCTAGTCCCTACTTTTAAACATATACAAAACCTTCTGCTTTTGTATAATA	73832	Db	72927	CTAAAGCAGTCTTGCCCGTGTGAGCTCTTGC-----	72898
QY	2098	CAATTTTCAATTAAGCCAGTATAAATAGAAACTAGCTTTTAAACTGAGCTTTGGAACCA	2157	QY	3168	CAGTCTTACTCCGATGTTTATCTTCTGATTTCAATGTGAATCCATCTTTCTGCAATATT	3227
Db	73831	TCAGGTTTATGAAGCCAGTAAAATAATAGGAACCTAGCTCTTCGATAAGCTTTGAAATCG	73772	Db	72897	-----TCAGTGGTATCCATCTTGGCTGCTGTTTT	72867
QY	2158	TTTGTGTTTCAAGTTTACCTGGGTATTTTGTCTGATTTCACTGCGCAATTTGCATTT	2217	QY	3228	TTGATGGAGAAATTTGAAGTTAAATGCAATTTAGAAACCTACCTGATGAACCTACCAAGT	3287
Db	73771	TTGCTGTTTAACTAGTTT-----GCTATTCTCTGATTTCTGCGAGTTGTCTAGT	73718	Db	72866	TGGATGGAGAACTTAACAATTAGACCCATCAGAAAGCACTTGAGGAGTT---CTACAGT	72810
				QY	3288	TTTAAAGTACTAGAAATATATACAGTAAATCCCACTTTCTATGATCTCTCGGAAATGAT	3347





REFERENCE  
AUTHORS

Rattus.  
1 (bases 1 to 278377)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbacia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisil,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,  
Homsai,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kurushi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,B., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,  
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,  
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telfrid,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs.R.  
Direct Submission  
Unpublished  
2 (bases 1 to 278377)  
Worley,K.C.  
Direct Submission

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (04-AUG-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 278377)  
Worley,K.C.  
Direct Submission

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Submitted (08-OCT-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Sep 19, 2002 this sequence version replaced gi:22538586.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented,  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

## COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: TUEG  
Center clone name: CH230-1B22

----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 220445 bases at least Q40  
Consensus quality: 225552 bases at least Q30  
Consensus quality: 228018 bases at least Q20  
Estimated insert size: 248009; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 247916: contig of 247916 bp in length  
247917 248016: gap of unknown length  
248017 272115: contig of 24099 bp in length  
272116 272215: gap of unknown length  
272216 273389: contig of 1374 bp in length  
273390 273689: gap of unknown length  
273690 275235: contig of 1546 bp in length  
275236 275335: gap of unknown length  
275336 276509: contig of 1174 bp in length  
276510 276609: gap of unknown length  
276610 278377: contig of 1768 bp in length.

FEATURES  
source

1..278377  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-1B22"

## misc\_feature

1..1312  
/note="wgs contig"

## misc\_feature

201866..203657  
/note="wgs contig"

## misc\_feature

246189..247916  
/note="wgs contig"

BASE COUNT 67642 a 48158 c 47319 g 67344 t 47914 others

## ORIGIN

Query Match 37.8%; Score 2761.2; DB 2; Length 278377;  
Best Local Similarity 72.8%; Pred. No. 0;  
Matches 4848; Conservative 0; Mismatches 1343; Indels 471; Gaps 81;  
QY 719 AAAGCAGATGGCTACTCATTTCTCTGAGAAAACCCACAAAGTGAATTAATTATTGTGAGAAA 778  
Db 144185 AAAGCAGATGGCTACTCATTTCTCTGAGAAAACCCACAAAGTGAATTAATTATTGTGAGAAA 144244  
QY 779 TGTTTACCGGAAGAGATCATCTGAGAAATCACTCCATACACAGACCTTAACAAAGAGA 838  
Db 144245 TGTTTACCGGAAGAGATCATCTGAGAAATCACTCCATACACAGACCTTAACAAAGAGA 144304  
QY 839 CGTTTAAGTCGAGAGATGTGGCAAGACTACATACCAAGCTTGGATTAAACCTCACT 898  
Db 144305 CTTTAAATGCGAGAGTGTGGCAAGAGCTACACCAAGCTTGGTTTAAGCGACACT 144364  
QY 899 TGGCTTTCATGCCGCAACAGTGTGTCACCTCACTGTAAGTATGTTTGCACACTTTTG 958  
Db 144365 TGGCTTTCATGCCGCAACAGTGTGTCACCTCACTGTAAGTATGTTTGCACACTTTTG 144424  
QY 959 AAAGCAGGGAGTGTCTCTGGAGCACCTTAATCTCATGAGGCAAGTGTCTGTGGGG 1018  
Db 144425 AGAGCAGAGTGTCTCTTAGAGCACCTGAATCTCAGCAGGCAAGTCTCTGGGGGG 144484  
QY 1019 TTAAGAAAAAAGACCAAGTCGAGCAATTTGTGATGCGCGGTTTCTACACCCGAAAGATG 1078

Db 144485 TGAAGAGAAAAGCACCAGTGTGACCACTGCGAGCGCAGGTTCTACACCCGGAAGATG 144544  
QY 1079 TCGGAGACACATGTTGGTGCACACTGGAAGAAGCACTTCCTCTCTCAGTATGTGCAC 1138  
Db 144545 TTGGGAGACACATGTTGGTGCACACGGGAAGAAGGACTTCCTCTCTCAGTATGTGCAC 144604  
QY 1139 AGAGATTGGGCGAAGAGATCCTGACCTCGACATATGAGAAGAGATCAATCAAGAGC 1198  
Db 144605 AGAGATTGGGCGAAGAGATCCTGACCTCGACATATGAGAAGAGATCAATCAAGAGC 144664  
QY 1199 TTCTGAAGTCAAAAACAGAACAGTGGATTTCTCTGACCCATTTACTCTGCAATGTCTCTG 1258  
Db 144665 TTCTGAAGTCAAAAACAGAACAGTGGATTTCTCTGACCCATTTCACTCTGCAATGTCTCTG 144724  
QY 1259 TGCTATAAAGACGAGCTCTCTTCGGTGTGATGCTCTTACCTCCAGTGAACCTGTTATCAA 1318  
Db 144725 TGCTATAAAGATGAGCTCTGCGGTCTATGTCCTTGGCTTCAGCGAAGCTCTTGTCAA 144784  
QY 1319 AGCCATTCAAAACACTTTCAGTTAAACCTCTACAACTCCATTTCACTGCTCCATCGACA 1378  
Db 144785 AGCCATTCAAAACACTTTCAGTTAAACCTCTATAACACTCCATTTCACTGCTCCATCGACA 144844  
QY 1379 GCTCGGATCTGCCACCAATGATCAAACTTTTACCTTTGGGAATGACATGCCCAATAG 1438  
Db 144845 GCTCTGGTCTACTCCAAATGATCAAACTTTTACCTTTGGGAATGACATGCCCAATAG 144904  
QY 1439 ATATGACACTGTTTCATCCCTCTCACCACTTTCTTCAATATCCGTTCACTTACCT 1498  
Db 144905 ATATGACACTGTTTCATCCCTCTCATCTCTGCTTCAATGCCGTTTCACTTACCT 144964  
QY 1499 CATATGCAATTTCTATCTCTGAAAGAACAGCCATTAAGGGGGAAATGAGAGTTACC 1558  
Db 144965 CATATGCAATTTCTATTTCCCGAGAAAGACAGCCCTTAAAGGGGGAAATGAGAGTTATC 145024  
QY 1559 TGATGAGTTTACAAAGTGGGCTGCTCTTCACTCCCAAGATTTCTCAAGCATCTGTCATCAT 1618  
Db 145025 TGATGAGTTTACAAAGTGGGCTGCTCTTCACTCCCAAGATTTCTCAAGCATCTGTCATCAT 145081  
QY 1619 CTAAGCTAGGTTGGATCTCTCAGATTTGGGTCCTTAGATGATGTTGAGGAGACCTCTCC 1678  
Db 145082 CTAAGTTAGGTTAGAGCTCTCAGATTTGGTCTCCAGAGATGTTGCTGGGGACCTCTCC 145141  
QY 1679 TATCCAAAGCTCTATCTCATCAGTACGACCCCTTAACACACAGCATTTGGATTTTCTC 1738  
Db 145142 TGTCAAAGAGCTCTATCTCTATCAGTACGACCCCTGAAACACACAGCATTTGGATTTTCTC 145201  
QY 1739 AGTTGTTAAATTTCAATGCTCTCTCTCTATAATCTCTATCAGTGGGAGCC 1798  
Db 145202 AGTTGTTCAATTTCAACATTAATGCTCTCTCTCTATCAGTGGGAGCC 145261  
QY 1799 TTGGAATGAGCTATTTCCGAGAGAAAGCAATTTCTCTGTTTCCAGCTCCCAACACAAA 1858  
Db 145262 TTGGAATGAGCTATTTCCGAGAGAAAGCAATTTCTCTCTCTATCAGTGGGAGCC 145321  
QY 1859 CACAGGATCTCAGGATCTGAAACACTATAGGCTTGGTCTCTGCACTCACTGTCAG 1918  
Db 145322 CACAGGATCTCAGGATCTGAAACACTATAGGCTTGGTCTCTGCACTCACTGTCAG 145381  
QY 1919 CAGCTTTCCAGCAGTTTAAAGCAACAGTACCACTCCCTCCCAAGCTTTCCATCAAGCTTTTC 1978  
Db 145382 CAGCTTTCCAGCAGCTTAAGCTCAAGCACTACCTGCTCCCGGTTTCCAGGCTTTTC 145441  
QY 1979 AGTAGGATCTGGGACATGATTCATTTACAGAAATGATGTAGTGTGCTGCTAGATGA 2038  
Db 145442 AGTAGGATCTGGGACATGATTCATTTACAGAAATGATGTAGTGTGCTGCTAGATGA 145499  
QY 2039 CCAATTTTATTTAGTGGCTTAAACAGATTAATTAATTTCTGCTTTGTATATAC 2098  
Db 145500 CCAATTTTATTTAGTGGCTTAAACAGATTAATTAATTTCTGCTTTGTATATAT 145559  
QY 2099 AAATTTTCAATAGCCAGTATAAATAGAACTAGCTTTTAAACTAGCTTTTGAACCAT 2158

Db 145560 CAGGTTTTATGAAGCCAGT---AAATAGGAATTAGCCTCTTTGATAAGCTTTGAAATGGT 145616  
QY 2159 TTGTGTTTCAAGTTAAAGTTTACCTGGGTTATTTTCTGCTGATTCTACTGCAATTTGTACATTT 2218  
Db 145617 TCGTGTGTTAACTACGTTT-----GCTGTTTCTCTGATTCTCTGCGAGTTGTCAATGTC 145670  
QY 2219 TAAGACTTTTTTTTTTTTCCATATAGGAAAGCCATTTATAGTAGTAACCTTTTACAAATCC 2278  
Db 145671 TCGGAGTTTATTT-----CATAGAGGAGGCCATTTATCTATAA---TTTACAAAGTCC 145725  
QY 2279 CATTTTCAAAATTTACTTTTATAGATCTTAAATTTTCTATTTTGTCTATAAATAGTGGCTCTA 2338  
Db 145726 TGTGTTCAAAATTTACTTGTAGACCTTAAATTTCTCACTTTTGTCTATAATAGCAATGG--- 145780  
QY 2339 CCTTTTGAATCTGCGCTCATTTAAATTTTAGCAATAGAAATGTAATTTGTATAAAGTT 2398  
Db 145781 GCATTTGCACTTTTTTTTTTTTGGCAATGTGAGTCAATGTAATCTCTCTAGAAATGTG 145840  
QY 2399 TGTGAATAAATCAAGGGTTT---AAATTTTCTTACTAGCTTCTTAAATGATTAATTAATCAA- 2456  
Db 145841 TGTGAAGAACTGAGTGTTCACATTTCTTACTAGCCAAATGATGATTAATCAATCAAG 145900  
QY 2457 -GTGCTTCAAAATGAAATTAAGAGTCCAGTTTCGGAAGATTAATGTTTGTAGATACAC 2515  
Db 145901 TGTGTTTAAATGAGTTAAGAAATCCAGTTTGGGGAGAGTATAGTGTTCCTTAAATACAC 145960  
QY 2516 CATATTTTCAAGTCACTATATTTCTGAAGACTCTCTGTTCTGCTGGCTTAAATATTTGCCAT 2575  
Db 145961 CATCATTTGAGATCAATATATTTTGAAGACT---GCTATTTGTCTGGCTTAAATAGAGTTT 146019  
QY 2576 CTTTATTTATGAGCTTTAAGAAAAACAAACCCCTTAAACACAAAGCATCAGTATTTATAGCA 2635  
Db 146020 CAGTA-----TCTATTT 146031  
QY 2636 AAAAGAGACTCTGTTAGGTGACATGTCATTTCTGTCATTAATAGTGTGGCCCTTAATTA 2695  
Db 146032 AATAGAAACCTAGAGATGTTACGGGTGTTCTGTTAG-----AACTA 146075  
QY 2696 GTACACAGGATATTTTGTGCTGTTTCTCATCTTCTTAAACATGCTATCTTTTCAATTAATA 2755  
Db 146076 GCCTTCAAGCTGATTTTGGCACATTTCAACATCTTTCCGT----- 146115  
QY 2756 TAGTAATAGTATGCTTGGGCTCTTCAGAGTTCGATATATAGTAGATCTCTTTAGTC 2815  
Db 146116 ---TACAGTGTGGCTCGGAGTCTTCAGGATCAGTGTAGAAGT---GTTCTGTGTATG 146171  
QY 2816 TTTTCCACCTTTTCAATCCAAAGGGGTGGGTCAAGTGCAGCAGCAATTTATTTTCTATGT 2875  
Db 146172 CCCTTCACTTTTATACCCCGAGGATGGCTTCA---CAATACAGCAATTTTACTTTGCATG 146229  
QY 2876 TGGCCACAGGTTAGTCCATTAATCTAGAGCCATTTGGAACCTGAGCCATCAGGTGTCTTT 2935  
Db 146230 TTGGTCCAGTTAGTGAGAGTCTTTAGAACATTTGGGACAAATAGACAGTAATATAGA 146289  
QY 2936 ATCCCAAGTGGATTTGACTCAGCTCTGT---GGGTGACAGACTTCTTAAGCAGGAAGATAG 2993  
Db 146290 AGCCCGAGTGGCTGGCTGCTGAATCTGCGGGTGTATAGCTCTTCCACAGGAAGATG 146349  
QY 2994 ACGTGAAGCACTGGTTTACATTTGGGAACCTTGTGTAGGGATCATGCGCCCTGTAGCCAGG 3053  
Db 146350 ATGTAAACCTGACAGCCAAACCCAGCACTTGTCTAGAGA---ATGTCCACTGTTGCCAGA 146407  
QY 3054 GTTAAACCTGAGCTTTTAAAGTAAGTAAAGCAATAGCGCTTATCATCTTCTTGTCT 3113  
Db 146408 ACTGAGAACTGGAGCTCTCTATCGCATCACTT----- 146437  
QY 3114 GAATTTGATATGTTTTTTTCTTCCCTTAAAGAAATCAAAAGCAGAAAAACAAACACAGTCC 3173  
Db 146438 -----CTAAACACAGCTCT 146452  
QY 3174 TACTCCGATGTTATCTTTCTGATTTCAATGTGAATCCATCTTCTCTGCAATATTTTGGAT 3233  
Db 146453 TGCC-----GTGCGCCCTTGCTCAGTGATATACATCTTCTCCTCGGTTTGGATG 146506



QY	3234	GGAGAAATTTGAAGTTAAATGCAATAGAAAACCTACCTGATGAACCTACCAAAAGTTTTTAAG	3293
Db	146507	GAGAACTTTAAATAATTAACACATCAAAAGCACTTGACGATT---CCACAGTTTACAA	146563
QY	3294	TGACTAGAAATATATACAGTAAATATCCACTTTTCATGCTCTCTGGGAAATGATAGAGT	3353
Db	146564	CCACAAAGTTG-----TTAAGAGCCACTTTTACACATCTCCAGGAGACAGAGGAAT	146618
QY	3354	ATTGCAATTAAGTTGAGTTTGTAGAGGTAAACAAAGTAAAGTAAACAAACCTATCTTGG	3413
Db	146619	GTTAGAAATTAAGTTGAAATTTATAGAGAAACAGAGGAAATAAGCAAAACCAAAATGAA	146678
QY	3414	TTAACATGAAAATAACAAATTTGAGAAATATATATATTTCACTGAATAATATATAGGCTTTTC	3473
Db	146679	TACGAATGCACCAAGGATGTGCTCACTGA-----GCAAGGTAGCAATGGGCTTCC	146731
QY	3474	TCACATTAGCAACCAACATAATCTCTTAAAGGTCTAAATATATATATTTTCTTAAGGT	3533
Db	146732	TCTTGTGACAGCAGCAACAAAGATGTCATCGTAGCT-CCGCGTATATCTTCTTAAGGT	146790
QY	3534	CAGTTGGGACATTAACCTTAGAACAATATCTATTAAGCACTTCTTAAACCTTATTTTAG	3593
Db	146791	TATTTGGGCACTCAACTAGACATTAC-----TAAGCACTTGTTAACACCTTATTTCTG	146846
QY	3594	GACCTTTTCGTTGGGATGGGGCAAGGTTGGGAGGTTTTTAAAGAGATATATATCTCT	3653
Db	146847	ACCCCTTCTGTTGGGCA-----GGAGTTTATAGAGATATATATCTCT	146893
QY	3654	TTAAAAAAAACAGAAAGAAA-----TATTTGACACTCAATTAGCCCT--A	3701
Db	146894	TTAAAAAAAACAGAAAGAAA-----TATTTGACACTCAATTAGCCCT--A	146953
QY	3702	TATGAAACTTCTTCTTCTTTTGTAGGCGCAGTTATCACTGCGAGATTGCAATGTTTACCA	3761
Db	146954	TGGGAGAGCTTCTTCTTGTGTAAGGCGGATGATCACTGCGATATGTTGTTTACC-147012	
QY	3762	AGAAATTTCTAAAAATGAGTGCAGATTACTCAATATATATATATTTTAAATATTTGGGA	3821
Db	147013	--CATCACTTAAGCAATAACAGATTACTGACTATGATGATTTTCAATGCGCTGGGA	147070
QY	3822	GTAGTAAATTTGTTGAGAAATGTAATTTGTAATATGTAATGTTGAGGCGCTTCAATATAT	3881
Db	147071	GTAGTCTAA---ATTGAGAAGTGAATCTGCTCT---GTGAATGGGCTTTCACATATAT	147123
QY	3882	ATATATAATACACACACACACATGACACATACACATACCCGCTTCAATGAATCAAGTTG	3941
Db	147124	GT-----CATACACAGCCTAGTGCACCTTCATAGAATCAGAGCTG	147163
QY	3942	CTCTCTGAGGAGCTTTGGCTCTCTGATATTTTATCATGCTCTCTATATTTTAAATCTTT	4001
Db	147164	CTCCCTAGAGG-GCTCAAGCTCTGTTCT-----AACCTT	147198
QY	4002	GGAGCAGTAGTTTTTATCTTATGTAATTTAAATTTTATGAAATTTTACATTTTAA	4061
Db	147199	AGAGCAGTAGTTTTTATACGTATGTAATTTAAATTTTATGTAATTTTAAATTTTAA	147257
QY	4062	AAAAGTGTCTCCAAAGCATTAATAATATATATGTTTAAATGAGGAGTACATTTTAAAT	4121
Db	147258	AAGAAAGTGTCTCCAAAGCATTAATAATATATATGTTTAAATGAGGAGTACATTTTAAAT	147315
QY	4122	TTTTCAAACTGCTCTCTAGCTTTTGTATGAGGAAATATTTTCTGAAAGTAGGCTTTTCG	4181
Db	147316	TCTTCAAGCTGCTCTGAGCTGAGGTTGGGCGGCACTCACTCAGAAAGTGGGCTCTTCG	147375
QY	4182	CTCTGCTTCACTAGCTTCTCTTATGTTTCTATGAAACAGATTTGCTTACCTAAATCTT-T	4240
Db	147376	CATGACCGC-----TTCCTTAGTTCTATGAAACCGATCACGCTACCTCTAAATGTTCT	147426
QY	4241	AGTTCAATGATGATGTTCAATATTTGCTTTTATACCA-----TATAAAGGAAAAAAT	4296
Db	147427	AGTTGGATGATGATGTTCAATATTTGCTTTTATACCACTTTTAAAGGAGGAGCAATTT	147486
QY	4297	GGTACAGACGACAAAT-----AGAAAACCTATTTTAAATAGAAATACCAAAATAGCAAG	4351
Db	147487	GGTACAGACGACAAAGTAAATAAAAAACCTATTTATTTATGTAGAGTTTACAGACACACA	147546
QY	4352	TGTGGAAGCACTACTTTATTTCTGTTTAAATGTACTTTAAAGGTTCATCAAAATTTAGTGAAC	4411
Db	147547	TATGAAAGCACTAGTTTATTTCTATCTAAAAATTTACTTTAAGAG---TCACATTTATGACC	147603
QY	4412	TGAGACATTTGGCTTTAGTAGGCTGTA-----TTCACTGTCTAAATTTAAAAAGGAGTACC	4466
Db	147604	TGAGACGTTGGCTTCTGAGCTGTATATATTTCACTGTCTAAATTTGAAAAAGGAGTACC	147663
QY	4467	AGCATTTATTAAGCTAAAGCACTTTTGGAAATGGGGAATAGCGGCATATATGTATGTATGTG	4526
Db	147664	AGCATTTATTAAGCTAAAGCACTTTTGGAAATGGGGAATAGCGGCATATATGTGTGTGTG	147721
QY	4527	TATGTGTGTGTGT-GGTGTGTGTATATATACACACACATACATCTTAAATCTTTCGCC	4585
Db	147722	TATATATATATATATATGTATGTATATATATATATATATATATATATATTTTTCCTAAT	147781
QY	4586	TGCATGAATTTCAATFACATGGAGGCACATCTTTCAGGGCACCACTGTATAAAATTTTGGAG	4645
Db	147782	TCTTGCCGAGCGAGAACTCTAATACACATCTTCCAGGGTACCACTGTATAAAGCTAGAG	147841
QY	4646	TCTTAATTT---TTTCATGTGTACACCTCTTTCCTGTCTCCACCCCGAGACT-TGAATAAA	4701
Db	147842	TCTTAATTTCTTGTGTGTATGTACACCTCTTTCCTGTCTCCACCCCGAGACTTGAATGA	147901
QY	4702	CACCTCAGAGTAAAGGGAATTCAGCTAATTT-TGTTTTTAAATTTCACTGTAGTGTCTAC	4760
Db	147902	CATTTCAAGTAAAGGGAATTAAGCAATTTCTTTTCGTTTATGACTGTAGTGTCTAC	147961
QY	4761	TAAACCTTTTTCAGAGAAATTTCTATTAAGATGAGGAGACTGCGTTATTTGAAATGCA	4820
Db	147962	TAAACCTTTTTCAGAGAAATTTCTATTAAGATGAGGAGACTGCGTTATTTGAAATGCA	148021
QY	4821	CAATGTTCTTAAAGGATGTAACAGAAATTTGCTTTTTCCTTA-----GAAAAGATTT	4877
Db	148022	CAATGTTTAAAGGATGTAACAGAAATTTGCTTTTTCCTTA-----GAAAAGATTT	148081
QY	4878	GTTTGTCTTCTAGTCAACTAGATATGATTAAATAAGATTTGCCAATCTGTTTTCATTT	4937
Db	148082	CACCTGTTCTGCTGAGTAGGACTGCTTAGCAATAGGCTTGTAGTGCCTTCTCTCATTT	148141
QY	4938	CTCTAGTGCACAGATTCATTTCTTGAATTTCTGTTAGTGCCTTGTGTTTAAAAA	4997
Db	148142	CTC-----CATGTTCTCTGGGATTTCTGATAGTGCCTTAGCTT-----148179	
QY	4998	AAAAAAAACCAAGGATTAACATTTAAATAAA---AGTAGTTTAGAATTTG	5054
Db	148180	-----AAAAAAGAGGTTTAAACATTTAACTAAACAGTGTAGTGTAGAGTTTG	148226
QY	5055	GGCTCTCAGACAGATTTGAACCTCAATTCAGTTTTCACATTTCCACATGATGTACAGTTAG	5114
Db	148227	GACCTCAGACAGATGCTGGGCGGCTTCAATTTTCACATTTCCACATGATGTGTGTGTGT	148282
QY	5115	GTACAAACAGGAGTTGAGTGTGGAAGGATCTTGGCACTGTGAAGCAATGCTATCCAT	5174
Db	148283	GCTTCAGGAGTGTAGCGCTTTTTCAGGAAAGGCTGTCGCACTGTGAGCAATATAGTCCAT	148342
QY	5175	TGATCTATACAGTACCTTTTATAGTTATCGATCAGT---TTAAACTTTTCAATTTTAAAT	5232
Db	148343	TGATATTTACCAATAATACCACTCTGTAGTTAATGTCTTCTTAAGCTTCCGTTTAAAT	148402
QY	5233	CCTATTACCAAGTTTCACTTTTAAACCTTCAATTTTCCTGCTGATTTATGCTACCTCT	5292
Db	148403	CCCGTTA-CTTGTGTAGGTTTCTAAATGTAAGTTGCTTGGAGATCATGACAGCTCT	148461
QY	5293	GTGTCAACTTTTTT-----ATTTCATTTAGTGTGTTTCTTTTCAAGCTGTGATTTT	5343
Db	148462	GTGTGAGCTTTTTTCCCTTTTAAATTTCACTGCGGTTTCTTTCAAGCTGTGTTGTGTTT	148521
QY	5344	GCCTATTGTTGCTTGTCTTTATTTTCTTAGTCAATTTTGTGGAAATATATAGTATATTTG	5403



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

# REFERENCE AUTHORS

1 (bases 1 to 244955)

Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, P., Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Louisedge, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankemeh, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, J., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C., Smajls, D., Sneed, A., Sodergren, E., Song, X., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabox, P., Taylor, C., Taylor, T., Thomas, R., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D., Holt, R., Smith, H., O., Weinstock, G., and Gibbs, R.A.

# TITLE JOURNAL

2 (bases 1 to 244955)

Worley, K.C.

# REFERENCE AUTHORS

Direct Submission

# TITLE JOURNAL

Submitted (29-May-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 244955)

Rat Genome Sequencing Consortium.

# REFERENCE AUTHORS

Direct Submission

# TITLE JOURNAL

Submitted (08-Oct-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 18, 2002 this sequence version replaced gi:21671673. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GRKE

Center clone name: CH230-208L11

Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 214612 bases at least Q40

Consensus quality: 217681 bases at least Q30

Consensus quality: 219976 bases at least Q20

Estimated insert size: 240534; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 5 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 135745: contig of 135745 bp in length

\* 135746 135845: gap of unknown length

\* 135846 239680: contig of 103835 bp in length

\* 239681 239780: gap of unknown length

\* 239781 240825: contig of 1045 bp in length

\* 240826 240925: gap of unknown length

\* 240926 242465: contig of 1540 bp in length

\* 242466 242566: gap of unknown length

\* 242566 244955: contig of 2390 bp in length.

Location/Qualifiers

1..244955

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-208L11"

BASE COUNT 63598 a 46725 c 47635 g 53647 t 23350 others

ORIGIN

Query Match 35.9%; Score 2623.8; DB 2; Length 244955;

Best Local Similarity 70.9%; Pred. No. 0;

Matches 4681; Conservative 0; Mismatches 1444; Indels 480; Gaps 79;

QY 796 CATCTGAAGATCATCTCCATACACACGACCTTAACCAAGAGACGTTTAAAGTGGAGAA 855

Db 223660 CATCTGAAGAACCACTCCATACACACGACCTTAACCAAGAGACGTTTAAAGTGGAGAA 223601

QY 856 TGTGCAAGAACTACAATACCAAGCTTGATTAAACGTCACCTTGGCCTTCATGCCGCA 915

Db 223600 TGTGCAAGAGCTACAACCAAGCTTGGGTTTAAAGCACAACCTTGGCCTTCATGCCGCTCC 223541

QY 916 ACAAGTGGTGACCTCACCTGTAAAGTATGTTTGCACCTTTGAAAGCACGGGAGTGCTT 975

Db 223540 ACCAGTGGCGACCTCACCTGTAAAGTATGTTTGCACCTTTGAGAGCACAGGTGTGCTC 223481

QY 976 CTGAGACACCTTAATCTCATGCGAGCAAGTCGTCTGTGGGGTTAAAGAAAAAAGCAC 1035

Db 223480 CTAGAGCACCTTGAATCTCATGCGAGCAAGTCGTCTGTGGGGTTAAAGAAAAAAGCAC 223421

QY 1036 CAGTGGCAACATCTGTATCGCCGCTTCTACACCCGAAGGATGTCGGGAGACACATGGTG 1095

Db	223420	CAGTGTGAGCACTGCGAGCGCAGGTTCTTACACCGGAAGGATGTTCCGAGACACATGGTG	223361	QY	2176	TAC-----CTGGTATTGTGCTCTGATTCACTGCGCAATTGTCACA	2215
QY	1096	CTGCACACTGGAGAGAGACTTCCTCTCTCAGTATTGTGCACAGAGATTGGCGAAG	1155	Db	222288	TGCTGTTTCNN	222229
Db	223360	GTGCACACGGGAAGAAGAGCTTCCTCTGTCTGAGTACTGTGCACAGAGATTGGACGAAG	223301	QY	2216	TTTTAAGACTTTTTTTTTTCCATATAGAAAGCCATTATTAGTAGTAACCTTTTACAA	2275
QY	1156	GATCAGCTGACCTGCACATATGAAGAAGAGTCAAAATCAAGAGCTTCTGAAGGTCAAAACA	1215	Db	222228	NN	222169
Db	223300	GATCAGCTCACTCGACACACATGAAGAAGAGTCAAAATCAAGAGCTTCTGAAGGTCAAAACT	223241	QY	2276	TCCCATTTTCAATTTACITTTTAGACTCTTAAATTTTCAATTTTGTCTTAATAACAGTGGCT	2335
QY	1216	GAACAGTGGATTTCTTGACCCCATTTACCTGCAATGTGTCTGTGCTATAAAGACGAG	1275	Db	222168	NN	222111
Db	223240	GAACAGTGGATTTCTTGACCCCATTTACCTGCAATGTGTCTGTGCTATAAAGATGAG	223181	QY	2336	CTACCTTTTGCATCTGGCTCATTAATAAATTTAGCAATAGAAATGTAAATTTGTATAAAA	2395
QY	1276	CTCCCTCCGGTGTATGCTTACCTTCCAGTGAAGTGTATCAAGAGCTTCTGAAGGTCAAAACT	1335	Db	222110	---GCATGCACTTTTTTTTTTTTGTGCAATGTGCAGTCAATGTAATCTCTATGAAAT	222054
Db	223180	CTCCCTCCGGTGTATGCTTACCTTCCAGTGAAGTGTATCAAGAGCTTCTGAAGGTCAAAACT	223121	QY	2396	GTGTGTGAAGAACTGAAGTGTTCACATTTCTTTACTAGCCACTAAATGGATTCATAATC	2454
QY	1336	TTGAGTTAAACCTCTACAACTCCATTTTCACTGATGAGAGCTCGGATCTGCCAC	1395	Db	222053	GTGTGTGAAGAACTGAAGTGTTCACATTTCTTTACTAGCCACTAAATGGATTCATAATC	221994
Db	223120	TTGAGTTAAACCTCTACAACTCCATTTTCACTGATGAGAGCTCGGATCTGCCAC	223061	QY	2455	AGTGC---TTCAAATGAATTAAGAGTCCAGTTTCGGAAGATATAAATGTTCTTAGATA	2512
QY	1396	CAATGATCAAACTTTACCTTTGGGAATGACATGCCCAATAGATGACACTGTTCAT	1455	Db	221993	AGTGTGTTTAAATGAGTTAAGAAATCCAGTTTGGGGAGATATAGTGGTTTCTTAATA	221934
Db	223060	CAATGATCAAACTTTACCTTTGGGAATGACATGCCCAATAGATGACACTGTTCAT	223001	QY	2513	CACATAAATTTCAAGTCAAGTATATCTGAAGACTCTCTGTGTCTGGCTAAAAATATTGC	2572
QY	1456	CCCTCTCAGCACCTTTCTTCAATATCCGTTTCAGTTCTACCTCATATGCAATTTCTATT	1515	Db	221933	CACCATCAATTTGAGATCAATATATTTTGAAGACT-GCTATTGTCTGCTAAATAAGAAG	221875
Db	223000	CCCTCTCAGCACCTTTCTTCAATATCCGTTTCAGTTCTACCTCATATGCAATTTCTATT	222941	QY	2573	CACTTTTATATGAGCTTTAAGGAAAAAACAACCCCTAAACAAGCAATCAGTATTATTA	2632
QY	1516	CTGTAAGAAAGAACAGAGCAATTAAGGGGGAATTTGAGAGTTACCTGATGGAGTTACAAGGT	1575	Db	221874	TTTCAGTA-----TCTA	221863
Db	222940	CCGAGAGAGAACAGAGCAATTAAGGGGGAATTTGAGAGTTACCTGATGGAGTTACAAGGT	222881	QY	2633	GCATAAGAGACTCTGTTAGTGCACATGGCAATTCGTGTCACTTAATAGTTGCGCCCTAAA	2692
QY	1576	GGCGTCCCTCTTTCATCCCAAGATTCTCAAGCATCGTCAATCACTAAGCTAGGGTTGGAT	1635	Db	221862	TTTAATAGAAACCTAGAGATGTACGGGGTGTCTGTGTAG-----AA	221819
Db	222880	GGCGCAACCATTTCTATCCCAAGATTCTCAAGCA---TCATCGTCTAAGTTAGGGTTAGAG	222824	QY	2693	TTAGTACACAGGATATTTGTGCTGTTCATCTTAAACATGCTATCTTTTCAITTTAA	2752
QY	1636	CTTCAGATTGGGTCCCTAGATGATGTGAGAGAGACTCTCCCTATCCAAAGCTCTATC	1695	Db	221818	CTAGCCCTCAGGCTGATTTGCCACATTTCAACCATCTTCCGT-----	221776
Db	222823	CTTCAGAGTGGTCCCGAGAGATGTGTGCGGAGCTCTCCCTGTCAAGAGCTCTATC	222764	QY	2753	TAATAGTAATAGTATGAGATGGGGTCTTCAAGTGCATATATAGTATAGTATCTCTTTA	2812
QY	1696	TCCATCAGTGACCCCTTAAACACACAGCAATTTGGATTTTCTCAGTTGTTTAAATTTCAAT	1755	Db	221775	-----TAACAGTGTGTGCGCTCGAGTCTTCCAGGATCAGTGTAGAAGT-GTTCTGTGT	221723
Db	222763	TCTATCAGTGACCCCTGAACACACAGCAATTTGATTTCTCTCAGTTGTTTCAATTTCAAT	222704	QY	2813	GTCTTTTCCACCTTTCACATCCAAAGGGTGGGTCAAGTGCAGCCAGCAATTTATTTTCAT	2872
QY	1756	CTTTTAAATGGTCCCTCTAATAATCTCTATCAGTGGGAGCTTTGGAATGAGCTATTC	1815	Db	221722	ATGCCCTTCACTTTTATACCCAGGATGGCTTCA--CAATACAGCATTTCTTACTTTGC	221665
Db	222703	CCATTAATGGTCCCTCTAATAATCTCTCTCGTGGGAGCTTTGGGATGAGCTATTC	222644	QY	2873	TGTTGGCCACAGGTTAGTCCATAATCTAGAGCCATTTGTGGAATCGAGCCATGAGGTGTG	2932
QY	1816	CAGGAGAGACATTTCTTCTGTTTCCAGCTCCCAACACACAGGATCTTCAGGAT	1875	Db	221664	ATGTTGTTCCAGTTAGTGCAGAGTCTTTAGAACCATTTGTGGGACAAATAGACAGTAATATA	221605
Db	222643	CAGGAGAGGCGCATTTCTCTGTTTCTCAGCTGCCACACAAACAGGATCTTCAGGAT	222584	QY	2933	TTTTATCCCAAGTGGATGACTCAGCCTCTGT--GGGTGACAGACTTCTAAGCAGGAAGA	2990
QY	1876	CTTGCAAACTATAGGGTGTGGGTCTCTGCATCTCACTGTGAGAGCTTTTCAACAGCAGT	1935	Db	221604	TGAAGCCAGAGTGGCTGCTGCTGAATCGTGGGGTGTATAGACTCTCCACAGGAAGA	221545
Db	222583	CTTGCGAACACAGTGGGCTTGGCTCTCTGCACTCACTGTGAGAGCTTTTCAACAGCAGC	222524	QY	2991	TAGACCTGAGACACATGGTTTACATTTTGGGAATTTGTAGGGATCATGGCCCTGTAGCC	3050
QY	1936	TGAACAGAGTACACCCCTCCAGCTTCCATCAAGCTTTTCAAGTAGGATCTGGACA	1995	Db	221544	TGGATGTAAACCTGACGCCAAACCCAGCATTGTCTAGAGA--ATGTCACCTGTGTC	221487
Db	222523	CTAAGCTCAGACTACCTCTGCCCGTTTCCACAGGCTTTTCAAGGCTCTGGGCGGA	222464	QY	3051	AGGGTTTAAAACTGGACTTTTAAAGTAAAGCAATAGCGCTTATATCATTTCTTT	3110
QY	1996	TGGATTCATTAAGAAATGATGTAGTGTGCTGCTAGATGACCATTTTATTTTATGTG	2055	Db	221486	AGAACTGAGAACTGGACTCTCTATCGCATCACTT-----	221454
Db	222463	GGGATTTCA--CCGGAGTGTCTGTGAGCCCCCGCTGCAAGACCATTTATATTTAGTG	222406	QY	3111	GCTGAATTTGATGATTTTTTCTTCCCTTAAGAAATCAAAAGCAGAAAAAACAACACACAG	3170
QY	2056	CCCTACTTTAAACAGATATAAAATTTCTGCTTTTGTATATAACAAATTTTCAATAGCCA	2115	Db	221453	-----CTAAACACACAG	221442
Db	222405	CCCTACTTTAAACAGACAGAACTTCTCCTTTTGTATATAATACAGGTTTATGAAGCCA	222346	QY	3171	TCCTACTCCGATGTTATCTTTCTGATTCATGTGAATCCATCTTCTTCTGCAATATTG	3230
QY	2116	GTATAAATAGAACTAGCTTTTAACTGAGCTTTGGAACCATTTGTGTTTCACTAGTT	2175	Db	221441	TCCTGCCC-----GTGCGCCCTTGTCTCAGTGCATATACATCTTCTCCCTGCCGTTTGG	221388
Db	222345	GT---AAATAGGAATTAGCCTCTTTGATAAGCTTTGAAATGGTTCGTTTAACTACGTT	222289				

[illegible]

[illegible]



Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1575)  
 Kas, K., Voz, M.L., Roijer, E., Astrom, A.K., Meyen, E., Stenman, G. and  
 Van de Ven, W.J.  
 Promoter swapping between the genes for a novel zinc finger protein  
 and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12)  
 translocations  
 Nat. Genet. 15 (2), 170-174 (1997)  
 97172974  
 9020842  
 2 (bases 1 to 1575)  
 Heusen, K., Voz, M.L., Van de Ven, W.J.M. and Kas, K.  
 Direct Submission  
 Submitted (06-APR-1998) Laboratory of Molecular Oncology, Center of  
 Human Genetics, Herestraat 49, Leuven 3000, Belgium  
 Location/Qualifiers  
 1..1575  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="129/SV"  
 /db\_xref="taxon:10090"  
 /map="t(3;8)(p21;q12)"  
 /tissue\_type="pleiomorphic adenoma"  
 1..1575  
 /gene="Plagl1"  
 /note="synonym: pleiomorphic adenoma gene 1"  
 21..1520  
 /gene="Plagl1"  
 /codon\_start=1  
 /product="C2H2 zinc finger protein PLAG1"  
 /protein\_id="AA21762.1"  
 /db\_xref="GI:6650195"  
 /translation="MATVIPGDLSEVRDTOKAPSKRRKRGSKRKNFPQCLDKAFN  
 SVEKLKHSFHTGERPKYKTHQDCTKAFVSKYKLRHMATHSPKTHKCNKCEKMF  
 RKDLKHLHTHDNKEFTFKCEECCKSYNTKLGFRHLAHLAASGDLTKVCQLQNF  
 STGLLEKSHAGSKSGGVEKKHQCEKRRFYTRKDVRRHVVHTLGRKDFLCQYC  
 AQRFGRKDLTRHMKSHQELLKVPKPEVPFDLPFTCNMSVPIKDELLPWSLPSSE  
 LLSKPFNTVNLNLYNTFQSMOSSGAHMLTTLPLGTCPIQMDAVHPSHHLAFKC  
 PFSSTVAISITPEKQPLKGEIYLMELQGGAPSSQSDSPASSKLGLEPQSGSPDD  
 GAGDLSKLSISISIDPLSTFALDFQNFIFPLNGPPYNTLSVGLSGMSYQBEAHS  
 SVSOLPTQDLPDANTVGLSSLSLSAFTSSLSSTTLPRFHAQ"  
 BASE COUNT 436 a 417 c 347 g 375 t  
 ORIGIN  
 Query Match 17.7%; Score 1295.6; DB 10; Length 1575;  
 Best Local Similarity 89.6%; Pred. No. 1.4e-220;  
 Matches 1416; Conservative 0; Mismatches 159; Indels 5; Gaps 2;  
 461 AAGGCTGGTTAGGCTGGATGGCCACTGTCAATTCCTGGTGAITTTGTCAGAACTAAGAG 520  
 1 AAGGCTGGTTAGGCTGGATGGCCACTGTCAATTCCTGGTGAITTTGTCAGAACTAAGAG 60  
 521 ATACCCAGAAAGTCCCTTCAGGAAACGTAAAGCGTGGTGAACCAACCAAGAAACAACT 580  
 61 ATACCCAGAAAGTCCCTTCAGGAAACGTAAAGCGTGGTGAACCAACCAAGAAACAACT 120  
 581 TTCCTTGCACACTGTGTGACAGGCTTTAAAGTGTGAGAAATTAAGGTTCACTCTCT 640  
 121 TTCCTTGCACACTGTGTGACAGGCTTTAAAGTGTGAGAAATTAAGGTTCACTCTCT 180  
 641 ACTCTCACACAGGAGAGGCGCTCAAGTGCATACAAACAGACTGCAACCAAGGCTTTG 700  
 181 TCTCTCACAGGAGAGGCGCTCAAGTGCATACAAACAGACTGCAACCAAGGCTTTG 240  
 701 TTTCTAAGTACAAATTAACAAAGGACATGGCTACTCTCTCTGAGAAACCAACCAAGT 760  
 241 TTTCTAAGTACAAATTAACAAAGGACATGGCTACTCTCTCTGAGAAACCAACCAAGT 300  
 761 GTAATTATTGTGAGAAATTTTACCGGAAGATCATCTGAAGATCACTCAATCATACAC 820  
 301 GTAATTATTGTGAGAAATTTTACCGGAAGATCATCTGAAGATCACTCAATCATACAC 360  
 821 ACGACCCCTAACAAAGAGACGTTTAAAGTGCAGAAATGTGGCAAGAACTACAATCCAC 880

Db 361 ATGACCCCAACAAAGAGACCTTTAAATGCAAGAGTGTGGCAAGACTACAATACCAAGC 420  
 QY 881 TTGGATTTAAAGCTCACCTTGGCTTGCATGCGCAACCAAGTGGTGCACCTCACTGTAAAG 940  
 Db 421 TTGGTTTTAAAGCGACACTTGGCTTGCATGCTGCCACAGTGGTGCACCTCACTGCAAG 480  
 QY 941 TATGTTTCAAACTTTTGAAGCAGCGGAGTGTCTCTTGAGAGCACTTAAATCTCATGCA 1000  
 Db 481 TGTGTTTGCAGAAATTTTGAAGCAGCGGAGTGTCTCTTGAGAGCACTTAAATCTCATG 540  
 QY 1001 GCAAGTCTGTGTCGGGTTTAAAGAAACCAAGTGGTGCAGCACTTGTGATCCCGGT 1060  
 Db 541 GCAAGTCTGTGTCGGGTTTAAAGAAACCAAGTGGTGCAGCACTTGTGATCCCGGT 600  
 QY 1061 TCTACACCCGAAAGGATGTCGGGAGACACATGCTGTCACACTGGAAGAAAGGACTTCC 1120  
 Db 601 TCTACACCCGAAAGGATGTCGGGAGACACATGCTGTCACACTGGAAGAAAGGACTTCC 660  
 QY 1121 TCTGTCAGTATTGTGCACAGAGATTGGGCGAAAGATCACTGACTCGACATATGAAGA 1180  
 Db 661 TCTGTCAGTATTGTGCACAGAGATTGGGCGAAAGATCACTGACTCGACATATGAAGA 720  
 QY 1181 AGAGTCACAATCAAGAGCTTCTGAAGGTCAAAACAGAACCAAGTGTCTTGTGACCCAT 1240  
 Db 721 AGAGTCACAATCAAGAGCTTCTGAAGGTCAAAACAGAACCAAGTGTCTTGTGACCCAT 780  
 QY 1241 TTACCTGCAATGTGTGCTGCTATAAAAGAGAGCTCTCTCCGGTGTGTCCTTACCTT 1300  
 Db 781 TTACCTGCAATGTGTGCTGCTATAAAAGATGAATCTCTGCCGTGATGTCCTTACCTT 840  
 QY 1301 CCAGTGAATGTATTCAAAAGCCATTCACAAACACTTTTGCAGTTAAACCTCTATAACACT 1360  
 Db 841 CCAGTGAATGTATTCAAAAGCCATTCACAAACACTTTTGCAGTTAAACCTCTATAACACT 900  
 QY 1361 CATTTCACTGTCATGAGAGCTCGGATCGCCATCCCAACCAATGATCACTTACCTTTGG 1420  
 Db 901 CATTTCACTGTCATGAGAGCTCTGGTCTGCTACCAATGATCACTTACCTTTGG 960  
 QY 1421 GAATGACATGCCAATAGATATGACACTGTTCATCCCTCTCACACCTCTTCTTCAAT 1480  
 Db 961 GAATGACATGCCAATAGATATGATGATGCTGTTCACCCCTCTCATCATCTTGTCTTCAAT 1020  
 QY 1481 ATCCGTTCACTGTTACCTCATATGCAATTTCTATTCTGAAAAAGAAAGCAAGCAATTAAG 1540  
 Db 1021 GCCCATTCAGTTCTACCTCATACGCAATCTCTATTCTGAAAAAGAAAGCAAGCAATTAAG 1080  
 QY 1541 GGGAAATTTGAGAGTTACCTGATGGAGTTACAGGTGGCGTCCCTCTTTCATCCCAAGATT 1600  
 Db 1081 GGGAAATTTGAGAGTTACCTGATGGAGTTACAGGTGGCGTCCCTCTTTCATCCCAAGATT 1140  
 QY 1601 CTCAGCATCTCATCATCTAAGCTTAGGTTGGATCTCTCAGATTGGGTCCTTAGATGATG 1660  
 Db 1141 CTCAGCATCTCTCTCTAAGCTTAGGTTGGATCTCTCAGATTGGGTCCTTAGATGATG 1197  
 QY 1661 GTGAGAGAGCTCTCTCTATCCAAAAGCTCTATCTCCATCAGTGAACCCCTTAAACACAC 1720  
 Db 1197 GTGAGAGAGCTCTCTCTATCCAAAAGCTCTATCTCCATCAGTGAACCCCTTAAACACAC 1257  
 QY 1721 CAGCATTTGGATTTTCTCAGTTGTTTAAATTTTACCTTTAAATGGTCTCTCTTATATC 1780  
 Db 1257 CAGCATTTGGATTTTCTCAGTTGTTTAAATTTTACCTTTAAATGGTCTCTCTTATATC 1317  
 QY 1781 CTCATCAGTGGGAGGCTTGGATGAGCTATTTCAGGAGAAAGAGCACTTCTCTGTTT 1840  
 Db 1317 CCCCTTTCAGTGGGAGGCTTGGATGAGCTATTTCAGGAGAAAGAGCACTTCTCTGTTT 1377  
 QY 1841 CCCAGCTTCCCAACAAACACAGGATCTTTCAGGATCTTGCAGAAACACTATAGGGCTTGG 1900  
 Db 1377 CTCAGCTTCCCAACAAACACAGGATCTTTCAGGATCTTGCAGAAACACTATAGGGCTTGG 1437  
 QY 1901 CTCCTGCACTCACTGTGACGAGCTTTTACCAGCAGTTTAAAGCAAGTACACCTTCCAC 1960



Db 1438 CTCTGCACTCACTGTGCGGAGCGTTTACCAGCAGCTGAGCTCAAGCACTTACTTGCCCC 1497

QY 1961 GTTTCATCAAGACTTTTTCAGTAGATCTGCGACATGGATTTCATTACAGAAATGATGTG 2020

Db 1498 GTTTCATCAAGCTTTTCAGTAGAGTCTGCGCGAGGAGTCA--CCCAGAGTGCTTGTG 1555

QY 2021 TAGCTGTGCGCTAGATGACC 2040

Db 1556 TAGCCCTGCGCTGCAATGACC 1575

RESULT 9

AC090187 65748 bp DNA linear HTG 13-APR-2001

LOCUS Homo sapiens chromosome 8 clone RP11-140116 map 8, LOW-PASS

DEFINITION SEQUENCE SAMPLING.

AC090187

VERSION AC090187.3 GI:13621271

KEYWORDS HTG; HTGS PHASE0.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE 1 (bases 1 to 65748)

JOURNAL Homo sapiens chromosome 8, clone RP11-140116

REFERENCE Unpublished

AUTHORS 2 (bases 1 to 65748)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glende, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, I., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Sougne, Z., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Apr 13, 2001 this sequence version replaced gi:13273404. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L12184

Center clone name: 140\_I\_16

NOTE: This record contains 78 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

742: contig of 742 bp in length

842: gap of 100 bp

843: contig of 766 bp in length

1609: gap of 100 bp

1709: contig of 727 bp in length

2436: gap of 100 bp

2536: contig of 703 bp in length

3339: gap of 100 bp

3339: contig of 716 bp in length

4055: gap of 100 bp

4155: contig of 742 bp in length

4897: gap of 100 bp

4997: contig of 761 bp in length

5758: gap of 100 bp

5858: contig of 759 bp in length

6617: gap of 100 bp

6717: contig of 773 bp in length

7489: gap of 100 bp

7589: contig of 765 bp in length

8355: gap of 100 bp

8455: contig of 763 bp in length

9218: gap of 100 bp

9318: contig of 722 bp in length

10039: gap of 100 bp

10140: contig of 729 bp in length

10868: gap of 100 bp

10969: contig of 732 bp in length

11701: gap of 100 bp

11800: contig of 753 bp in length

12553: gap of 100 bp

12554: contig of 786 bp in length

13439: gap of 100 bp

13440: contig of 765 bp in length

14304: gap of 100 bp

14405: contig of 748 bp in length

15152: gap of 100 bp

15153: contig of 764 bp in length

16016: gap of 100 bp

16117: contig of 746 bp in length

16862: gap of 100 bp

16863: contig of 723 bp in length

17685: gap of 100 bp

17686: contig of 727 bp in length

17786: gap of 100 bp

18513: contig of 729 bp in length

19341: gap of 100 bp

19342: contig of 734 bp in length

20175: gap of 100 bp

20276: contig of 739 bp in length

21014: gap of 100 bp

21115: contig of 765 bp in length

21879: gap of 100 bp

21880: contig of 757 bp in length

22737: gap of 100 bp

22837: contig of 780 bp in length

23616: gap of 100 bp

23717: contig of 699 bp in length

24415: gap of 100 bp

24516: contig of 770 bp in length

25285: gap of 100 bp

25386: contig of 774 bp in length

26159: gap of 100 bp

26260: contig of 773 bp in length

27032: gap of 100 bp

27133: contig of 720 bp in length

27852: gap of 100 bp

27953: contig of 734 bp in length

28686: gap of 100 bp

28687: contig of 722 bp in length

29508: gap of 100 bp

29509: contig of 100 bp

Query Match 9.7%; Score 707.4; DB 2; Length 65748;  
Best Local Similarity 86.6%; Pred. No. 4e-116;  
Matches 711; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 5611 TAGCACCTTTCAPAAATGNAATCTAAACCTTTACAGAAACTTCTCTATAGGTATAGCCTTTTA 5670

Db	53982	TGCGATTCTCAATGNAATCTAAACCTTTTACAGAACTTCTTATAGGTATGCTTTTA	54041
QY	5671	TTTTAAGACTTATATAATTCAAAGTGCCATTAGATGATATATATGTAGGCCCTTTGATATA	5730
Db	54042	TTTTAAGACTTATATAATTCAAAGTGCCATTAGATGATATATATGTAGGCCCTTTGATATA	54101
QY	5731	TAATGCTTTGTGTACAAAATGCTAGATGGTATTTTTAAACAGGTACACATTTTACAGTGT	5790
Db	54102	TAATGCTTTGTGTACAAAATGCTAGATGGTATTTTTAAACAGGTACACATTTTACAGTGT	54161
QY	5791	TTCTTATCAATTTGCTATATTGCACAGAAATCAGTGTGTCTTTTCATAAGGTTTTACAA	5850
Db	54162	TTCTTATCAATTTGCTATATTGCACAGAAATCAGTGTGTCTTTTCATAAGGTTTTACAA	54221
QY	5851	TGGTTATTTTTTTTACAAAGTTTACGTGTCTCAAAGCACACTGTCTTCCCAGTACGTAAG	5910
Db	54222	TGGTTATTTTTTTTACAAAGTTTACGTGTCTCAAAGCACACTGTCTTCCCAGTACGTAAG	54281
QY	5911	TTAAAAATATCACAGTTTACCCAAAGTTGCTTACGCCCTACTGAGATCCATGTGACATTTGA	5970
Db	54282	TTAAAAATATCACAGTTTACCCAAAGTTGCTTACGCCCTACTGAGATCCATGTGACATTTGA	54341
QY	5971	GGAGATCTTTTAAAATGTTTATGATTTCTGTCATTTAGCAATGGCTGGCTGTAGTTCTGGTAA	6030
Db	54342	GGAGATCTTTTAAAATGTTTATGATTTCTGTCATTTAGCAATGGCTGGCTGTAGTTCTGGTAA	54401
QY	6031	ATGTGTGCTAAGTTGAATTTGCTGTGTTTCTCACACTGTGTACAGCCATGTCTAC	6090
Db	54402	ATGTGTGCTAAGTTGAATTTGCTGTGTTTCTCACACTGTGTACAGCCATGTCTAC	54461
QY	6091	AACACAGATAAGTCTGTTGTGATCATATAGATCTACATAAGTTGTGCAGTTTGTGCTAA	6150
Db	54462	AACACAGATAAGTCTGTTGTGATCATATAGATCTACATAAGTTGTGCAGTTTGTGCTAA	54521
QY	6151	AAACCCATAGGAGAGCTCCTTTGGGATCATAGAAAAGAGATCATGCAACCCAGCATTTGGTG	6210
Db	54522	AAACCCATAGGAGAGCTCCTTTGGGATCATAGAAAAGAGATCATGCAACCCAGCATTTGGTG	54581
QY	6211	AAGGCACACTCAGATTCAGCTTAGGGCTTTCTATGATTTGTTCACACCTCTGAGGATGG	6270
Db	54582	AAGGCACACTCAGATTCAGCTTAGGGCTTTCTATGATTTGTTCACACCTCTGAGGATGG	54641
QY	6271	AAGGCAGTGTCTTTTGATGTTATCTAGCCTGAAATGACACAGACTATTGCTAATGTAT	6330
Db	54642	AAGGCAGTGTCTTTTGATGTTATCTAGCCTGAAATGACACAGACTATTGCTAATGTAN	54701
QY	6331	AAAAACATTCATATATAAGCTTCAGTGGTACAGATGAACCAAGATGAATTTATCTTC	6390
Db	54702	NN	54761
QY	6391	TCAGAAACACTCCTTCAATATATATTGGATCATGCTGCTA	6431
Db	54762	NN	54802
RESULT 10			
AC090187/c			
LOCUS	AC090187	65748 bp	DNA
DEFINITION	Homo sapiens chromosome 8 clone RP11-140116 map 8, LOW-PASS		
SEQUENCE SAMPLING	linear		
ACCESSION	AC090187		
VERSION	AC090187.3	GI:13621271	
KEYWORDS	HTG; HTGS_PHASE0.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 65748)		
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
JOURNAL	Homo sapiens chromosome 8, clone RP11-140116		
REFERENCE	2 (bases 1 to 65748)		

## AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Boguslavskiy, L., Bouckhgalter, B., Brown, A.,  
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T.,  
 Lehoczyk, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,  
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,  
 McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mienga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,  
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,  
 Roy, A., Santos, R., Schauer, S., Stange-Thomann, N., Stojanovic, N.,  
 Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Apr 13, 2001 this sequence version replaced gi:13273404.

All repeats were identified using RepeatMasker:

Smith, A.P.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L12184

Center clone name: 140\_I\_16

-----  
 \* NOTE: This record contains 78 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 742: contig of 742 bp in length  
 \* 743 842: gap of 100 bp  
 \* 843 1608: contig of 766 bp in length  
 \* 1609 1708: gap of 100 bp  
 \* 1709 2435: contig of 727 bp in length  
 \* 2436 2535: gap of 100 bp  
 \* 2536 3238: contig of 703 bp in length  
 \* 3239 3338: gap of 100 bp  
 \* 3339 4054: contig of 716 bp in length  
 \* 4055 4155: gap of 100 bp  
 \* 4156 4896: contig of 742 bp in length  
 \* 4897 4996: gap of 100 bp  
 \* 4997 5757: contig of 761 bp in length  
 \* 5758 5857: gap of 100 bp  
 \* 5858 6616: contig of 759 bp in length  
 \* 6617 6716: gap of 100 bp  
 \* 6717 7489: contig of 773 bp in length  
 \* 7490 7590: gap of 100 bp  
 \* 7591 8354: contig of 765 bp in length  
 \* 8355 8454: gap of 100 bp  
 \* 8455 9217: contig of 763 bp in length  
 \* 9218 9317: gap of 100 bp  
 \* 9318 10039: contig of 722 bp in length  
 \* 10040 10139: gap of 100 bp  
 \* 10140 10868: contig of 729 bp in length

10869 10968: gap of 100 bp  
 \* 10969 11700: contig of 732 bp in length  
 \* 11701 11800: gap of 100 bp  
 \* 11801 12553: contig of 753 bp in length  
 \* 12554 12553: gap of 100 bp  
 \* 12554 13439: contig of 786 bp in length  
 \* 12554 13440: gap of 100 bp  
 \* 13440 14304: contig of 765 bp in length  
 \* 14304 14404: gap of 100 bp  
 \* 14405 15152: contig of 748 bp in length  
 \* 15152 15352: gap of 100 bp  
 \* 15353 16016: contig of 764 bp in length  
 \* 16017 16116: gap of 100 bp  
 \* 16117 16862: contig of 746 bp in length  
 \* 16863 16962: gap of 100 bp  
 \* 16963 17685: contig of 723 bp in length  
 \* 17686 17785: gap of 100 bp  
 \* 17786 18512: contig of 727 bp in length  
 \* 18513 18612: gap of 100 bp  
 \* 18613 19341: contig of 729 bp in length  
 \* 19341 19441: gap of 100 bp  
 \* 19442 20175: contig of 734 bp in length  
 \* 20176 20275: gap of 100 bp  
 \* 20276 21014: contig of 739 bp in length  
 \* 21015 21114: gap of 100 bp  
 \* 21115 21879: contig of 765 bp in length  
 \* 21880 21979: gap of 100 bp  
 \* 21980 22736: contig of 757 bp in length  
 \* 22737 22837: gap of 100 bp  
 \* 22837 23616: contig of 780 bp in length  
 \* 23617 23716: gap of 100 bp  
 \* 23717 24415: contig of 699 bp in length  
 \* 24416 24515: gap of 100 bp  
 \* 24516 25285: contig of 770 bp in length  
 \* 25286 25386: gap of 100 bp  
 \* 25386 26159: contig of 774 bp in length  
 \* 26160 26259: gap of 100 bp  
 \* 26260 27032: contig of 773 bp in length  
 \* 27033 27132: gap of 100 bp  
 \* 27133 27852: contig of 720 bp in length  
 \* 27853 27953: gap of 100 bp  
 \* 27953 28686: contig of 734 bp in length  
 \* 28687 28786: gap of 100 bp  
 \* 28787 29508: contig of 722 bp in length  
 \* 29509 29608: gap of 100 bp  
 \* 29609 30359: contig of 751 bp in length  
 \* 30360 30459: gap of 100 bp  
 \* 30460 31206: contig of 747 bp in length  
 \* 31207 31306: gap of 100 bp  
 \* 31307 32050: contig of 744 bp in length  
 \* 32051 32150: gap of 100 bp  
 \* 32151 32912: contig of 762 bp in length  
 \* 32913 33012: gap of 100 bp  
 \* 33013 33739: contig of 727 bp in length  
 \* 33740 33839: gap of 100 bp  
 \* 33840 34580: contig of 741 bp in length  
 \* 34581 34680: gap of 100 bp  
 \* 34681 35373: contig of 693 bp in length  
 \* 35374 35473: gap of 100 bp  
 \* 35474 36191: contig of 718 bp in length  
 \* 36192 36291: gap of 100 bp  
 \* 36292 37022: contig of 731 bp in length  
 \* 37023 37122: gap of 100 bp  
 \* 37123 37874: contig of 752 bp in length  
 \* 37875 37974: gap of 100 bp  
 \* 37975 38731: contig of 757 bp in length  
 \* 38732 38831: gap of 100 bp  
 \* 38832 39605: contig of 774 bp in length  
 \* 39606 39705: gap of 100 bp  
 \* 39706 40465: contig of 760 bp in length  
 \* 40466 40565: gap of 100 bp  
 \* 40566 41271: contig of 706 bp in length  
 \* 41272 41371: gap of 100 bp



----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: L24504  
 Center clone name: 1053\_A\_18

\* NOTE: This record contains 62 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 693 592: contig of 692 bp in length  
 792: gap of 100 bp  
 793 1456: contig of 664 bp in length  
 1457 1556: gap of 100 bp  
 1557 2165: contig of 609 bp in length  
 2166 2285: gap of 100 bp  
 2286 2963: contig of 698 bp in length  
 2964 3063: gap of 100 bp  
 3064 3778: contig of 715 bp in length  
 3779 3878: gap of 100 bp  
 3879 4538: contig of 660 bp in length  
 4539 4638: gap of 100 bp  
 4639 5277: contig of 639 bp in length  
 5278 5377: gap of 100 bp  
 5378 6099: contig of 722 bp in length  
 6100 6199: gap of 100 bp  
 6199 6892: contig of 693 bp in length  
 6893 6992: gap of 100 bp  
 6993 7707: contig of 715 bp in length  
 7708 7807: gap of 100 bp  
 7808 8515: contig of 708 bp in length  
 8516 9299: gap of 100 bp  
 9299 9399: contig of 684 bp in length  
 9399 9400 9399: gap of 100 bp  
 10101: contig of 702 bp in length  
 10201: gap of 100 bp  
 10202 10896: contig of 695 bp in length  
 10897 10996: gap of 100 bp  
 10997 11660: contig of 664 bp in length  
 11661 11760: gap of 100 bp  
 11761 12438: contig of 678 bp in length  
 12439 12538: gap of 100 bp  
 12539 13233: contig of 695 bp in length  
 13234 13333: gap of 100 bp  
 13334 14057: contig of 724 bp in length  
 14058 14157: gap of 100 bp  
 14158 14838: contig of 681 bp in length  
 14839 14938: gap of 100 bp  
 14939 15642: contig of 704 bp in length  
 15643 15742: gap of 100 bp  
 15743 16461: contig of 719 bp in length  
 16462 16561: gap of 100 bp  
 16562 17250: contig of 689 bp in length  
 17251 17350: gap of 100 bp  
 17351 18053: contig of 703 bp in length  
 18054 18153: gap of 100 bp  
 18154 18862: contig of 709 bp in length  
 18863 18962: gap of 100 bp  
 18963 19683: contig of 721 bp in length  
 19684 19784: gap of 100 bp  
 19784 20488: contig of 705 bp in length  
 20489 20588: gap of 100 bp  
 20589 21312: contig of 724 bp in length

21313 21412: gap of 100 bp  
 21413 22101: contig of 689 bp in length  
 22102 22201: gap of 100 bp  
 22202 22905: contig of 704 bp in length  
 22906 23005: gap of 100 bp  
 23006 23707: contig of 702 bp in length  
 23708 23807: gap of 100 bp  
 23808 24498: contig of 691 bp in length  
 24499 24598: gap of 100 bp  
 24599 25104: contig of 706 bp in length  
 25105 25404: gap of 100 bp  
 25405 26090: contig of 686 bp in length  
 26091 26190: gap of 100 bp  
 26191 26895: contig of 705 bp in length  
 26896 26995: gap of 100 bp  
 26996 27704: contig of 709 bp in length  
 27705 27804: gap of 100 bp  
 27805 28495: contig of 691 bp in length  
 28496 28595: gap of 100 bp  
 28596 29294: contig of 699 bp in length  
 29295 29394: gap of 100 bp  
 29395 30081: contig of 687 bp in length  
 30082 30181: gap of 100 bp  
 30182 30859: contig of 678 bp in length  
 30860 30959: gap of 100 bp  
 30960 31677: contig of 718 bp in length  
 31678 31777: gap of 100 bp  
 31778 32492: contig of 715 bp in length  
 32493 32592: gap of 100 bp  
 32593 33300: contig of 708 bp in length  
 33301 33400: gap of 100 bp  
 33401 34097: contig of 697 bp in length  
 34098 34197: gap of 100 bp  
 34198 34878: contig of 681 bp in length  
 34879 34978: gap of 100 bp  
 34979 35692: contig of 714 bp in length  
 35693 35792: gap of 100 bp  
 35793 36515: contig of 723 bp in length  
 36516 37333: contig of 718 bp in length  
 37334 37433: gap of 100 bp  
 37434 38135: contig of 702 bp in length  
 38136 38235: gap of 100 bp  
 38236 38935: contig of 700 bp in length  
 38936 39035: gap of 100 bp  
 39036 39740: contig of 705 bp in length  
 39741 39840: gap of 100 bp  
 39841 40557: contig of 717 bp in length  
 40558 40657: gap of 100 bp  
 40658 41353: contig of 696 bp in length  
 41354 42181: contig of 728 bp in length  
 42182 42281: gap of 100 bp  
 42282 42942: contig of 661 bp in length  
 42943 43042: gap of 100 bp  
 43043 43772: contig of 730 bp in length  
 43773 43872: gap of 100 bp  
 43873 44532: contig of 660 bp in length  
 44533 44632: gap of 100 bp  
 44633 45343: contig of 711 bp in length  
 45344 45443: gap of 100 bp  
 45444 46167: contig of 724 bp in length  
 46168 46267: gap of 100 bp  
 46268 46986: contig of 719 bp in length  
 46987 47086: gap of 100 bp  
 47087 47788: contig of 702 bp in length  
 47789 47888: gap of 100 bp  
 47889 48549: contig of 661 bp in length  
 48550 48649: gap of 100 bp  
 48650 49357: contig of 708 bp in length.

Location/Qualifiers

1. .49357  
 /organism="Homo sapiens"

FEATURES  
 source

```
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-1053A18"
/clone_lib="RP11-11 Human Male BAC"
BASE COUNT 12503 a 9244 c 8714 g 12754 t 6142 others
ORIGIN

Query Match      8.8%; Score 645.4; DB 2; Length 49357;
Best Local Similarity 99.8%; Pred. No. 4.6e-105;
Matches 646; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5155 CTGTAAGCAATGCTATCCATTCATGATATACCAAGTACCTTTATAGTTATCGATCAGCTGTTA 5214
Db 30213 CTGTAAGCAATGCTATCCATTCATGATATACCAAGTACCTTTATAGTTATCGATCAGCTGTTA 30272
QY 5215 AAACCTTCATTTAAATCCTATTAACCAAGTTCAGTTTTTAAACCTTCAATGTCCTGG 5274
Db 30273 AAACCTTCATTTAAATCCTATTAACCAAGTTCAGTTTTTAAACCTTCAATGTCCTGG 30332
QY 5275 CTGATTATGCATCAGCTCTGTGTGCACTTTTATATTCATTTAGTTTCTTTCAGCTG 5334
Db 30333 CTGATTATGCATCAGCTCTGTGTGCACTTTTATATTCATTTAGTTTCTTTCAGCTG 30392
QY 5335 TGTATTTTGGCCTATTTGTTGCTGTGCTTTATTTTCTTCTAGTTCATTTGTCGAATATAGT 5394
Db 30393 TGTATTTTGGCCTATTTGTTGCTGTGCTTTATTTTCTTCTAGTTCATTTGTCGAATATAGT 30452
QY 5395 GATATATTGTGTTAATTGCGACAGTAGCGGTTTTTAAACCAATATATCTAGTGAACAT 5454
Db 30453 GATATATTGTGTTAATTGCGACAGTAGCGGTTTTTAAACCAATATATCTAGTGAACAT 30512
QY 5455 GAGCCAGAGCGGATTCCTTTATTAAGCTTAATATGAATGTTAAAGAGTACATATTTTCAG 5514
Db 30513 GAGCCAGAGCGGATTCCTTTATTAAGCTTAATATGAATGTTAAAGAGTACATATTTTCAG 30572
QY 5515 GATCGTTTCATCTAGTGCAGCAATACATATTAAGCCCAATATTTTAAACCAATAGAG 5574
Db 30573 GATCGTTTCATCTAGTGCAGCAATACATATTAAGCCCAATATTTTAAACCAATAGAG 30632
QY 5575 CTGTGTCACCTCTATACATATTAACAGATATAGCACTTTCAAATGAATCTAAA 5634
Db 30633 CTGTGTCACCTCTATACATATTAACAGATATAGCACTTTCAAATGAATCTAAA 30692
QY 5635 CTTTACAGAACTTCTTATAGGTTATGCCCTTTATTTTAAAGCTTATTAATTAATCAAG 5694
Db 30693 CTTTACAGAACTTCTTATAGGTTATGCCCTTTATTTTAAAGCTTATTAATTAATCAAG 30752
QY 5695 TGCCATTAGATGATATATATAGGCTTTGATATATATATGCTTTGTGTACAAAAATCGT 5754
Db 30753 TGCCATTAGATGATATATATAGGCTTTGATATATATATGCTTTGTGTACAAAAATCGT 30812
QY 5755 AGATGGTATTTTAAACAGGTACATTTTACAGTGTGTTTCTTATCAAT 5801
Db 30813 AGATGGTATTTTAAACAGGTACATTTTACAGTGTGTTTCTTATCAAT 30859

RESULT 12
LOCUS G25280/c 421 bp DNA linear STS 02-JUN-1996
DEFINITION human STS EST334260, sequence tagged site.
ACCESSION G25280
VERSION G25280.1 GI:1347512
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Hudson,I.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STSs
```

## JOURNAL COMMENT

Unpublished (1995)

Contact: Thomas Hudson  
Whitehead Institute/MIT Center for Genome Research  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge MA 02142 USA  
Tel: 617 252 1900  
Fax: 617 252 1902  
Email: thudson@genome.wi.mit.edu

Primer A: AACCAAGCTAAGGCACCTACCA  
Primer B: CACAGAATTGGCTTTTTCCTC  
STS size: 150  
PCR Profile:

Presoak:  
Denaturation:  
Annealing: 56 degrees C  
Polymerization:  
PCR Cycles: 35  
Thermal Cycler:

Protocol:  
Template: 10 ng  
Primer: each 5 pm  
dNTPs: each 4 mM  
Tag Polymerase: 0.025 units/ul  
Total Vol: 20 ul

## Buffer:

MgCl2: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 9.3

Derived from dbEST (genbank accession D59273).

## FEATURES

Location/Qualifiers  
1..421  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="314.2 cR from top of Chr8 linkage group"

## STS

primer\_bind 1..21

primer\_bind complement(129..150)

BASE COUNT 144 a 77 c 75 g 122 t 3 others

## ORIGIN

Query Match 5.7%; Score 416; DB 11; Length 421;  
Best Local Similarity 99.5%; Pred. No. 7.3e-64;  
Matches 416; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4575 AAATCTTCCCTGCATGAAATCAAATACATGGAGGCACATCTTCAGGGCACCAGTGTGA 4634  
Db 418 AAANCTTGCCTCCGATGAAATCAAATACATGGAGGCACATCTTCAGGGCACCAGTGTGA 359  
QY 4635 AAATTTTGGAGTCTTAATTTTCATGTGTACACCTCTTTGCCCTGTCCACCCCGAGCTT 4694  
Db 358 AAATTTTGGAGTCTTAATTTTCATGTGTACACCTCTTTGCCCTGTCCACCCCGAGCTT 299  
QY 4695 GAAATAACACCTTCAGAGTAAGAGGGAATTCAGCTAAATTTGTTTTTAAATTTGACTGTAGT 4754  
Db 298 GAAATAACACCTTCAGAGTAAGAGGGAATTCAGCTAAATTTGTTTTTAAATTTGACTGTAGT 239  
QY 4755 GGTCACTAAACCCCTTTTGGAGAGAAATTTCTATTAAAGATGAGGAGACTCGCTTATTGGA 4814  
Db 238 GGTCACTAAACCCCTTTTGGAGAGAAATTTCTATTAAAGATGAGGAGACTCGCTTATTGGA 179  
QY 4815 ATTGCACAATGTTCTTAACAGAGATGTAAACAGAAATTTGGCTTTTTTTCCTAGAAAAAG 4874  
Db 178 ATTGCACAATGTTCTTAACAGAGATGTAAACAGAAATTTGGCTTTTTTTCCTAGAAAAAG 119  
QY 4875 ATTGTTTGTCTTCTATGTCACTAGATATGATTAATAAATAAGTATTCGAATGCTGTTTTC 4934  
Db 118 ATTGTTTGTCTTCTATGTCACTAGATATGATTAATAAATAAGTATTCGAATGCTGTTTTC 59

QY	4935	ATTCTCTAGTGGCCAGAAATCATTTCTCTGTAATTTCTGTAGTGCCTTAGCTTGTT	4992
Db	58	ATTCTCTAGTGGCCAGAAATCATTTCTCTGTAATTTCTGTAGTGCCTTAGCTTGTT	1
RESULT 13			
LOCUS	BC023655	1829 bp mRNA linear PRI 23-SEP-2002	
DEFINITION	Homo sapiens, clone MGC:29597 IMAGE:4764127, mRNA, complete cds.		
ACCESSION	BC023655		
VERSION	BC023655.1	GI:23271119	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1829)		
JOURNAL	Straussberg, R.		
	Direct Submission		
	Submitted (05-FEB-2002) National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk		
	Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a>		
	Tissue Procurement: Louis Staudt		
	cDNA Library Preparation: Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: National Institutes of Health Intramural		
	Sequencing Center (NISC),		
	Gaithersburg, Maryland;		
	Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a>		
	Contact: <a href="mailto:nisc.mgc@nih.gov">nisc.mgc@nih.gov</a>		
	Akhter, N., Ayelle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,		
	Blakesley, R.W., Bouffard, G., Green, K., Brinkley, C., Brooks, S.,		
	Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,		
	Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,		
	Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, P.D., McCloskey, J.C.,		
	McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,		
	Tsurgou, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,		
	Young, A., Zhang, L.-H. and Green, E.D.		
	Clone distribution: MGC clone distribution information can be found		
	through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Series: IRAL Plate: 39 Row: f Column: 4		
	This clone was selected for full length sequencing because it		
	passed the following selection criteria: matched mRNA gi: 6031195.		
FEATURES	Location/Qualifiers		
source	1. .1829		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="MGC:29597 IMAGE:4764127"		
	/tissue_type="Primary B-Cells from Tonsils"		
	/clone_lib="NIH MGC 48"		
	/lab_host="DH10B-R"		
	/note="Vector: pOTB7"		
	213. .1703		
	/codon_start=1		
	/product="Unknown (protein for MGC:29597)"		
	/protein_id="AAH23655.1"		
	/db_xref="GI:23271120"		
	/translation="MTFTFTSVPPWLTQDAKQEEVGVKLVPRPRGAEASOVKCCCEI		
	SGTFFSNGEKLRPHSLUPQEPQRYSCPLHCGKAPASKYKLYRHMATHSAQKPHQMY		
	CDKMFHRKDLNHLQTHDPNKEALHCBGKNYNTKLGYYRHLRHLAAHSSVGLSCKV		
	CLQTFSTQALLHLKHAHRVAGAKKKPCDCHDRFRYTRKDYRRLVHLVHTGRK		
	FLCOYCAORFGRKDLHTRVKKSHQELLKITEPVDMLGLLSCSTVSVKELSPVL		
	CMASRDVMTGKAPGMLPMCMYGAHITPMTSGVPHSLVHNTLPMGMSVPLESSPTSS		
	PAOLPKYVOLGSYLPDLKPKVVDSEFLAEPLGSLSLSAEPQAPASPOPAALALID		
	EALLAKSPANLSALCAANVDFSHLGLFPLNLNPPCNPPGATGLVNGYSQAEPALL		
	TTTQAQPDQDSFGAGGGLNFGPLHSLPFPVFTSGLSLSTTLPRFHOAFQ"		

QY	577	AACTTTCTTCCTGCAACTGTGTGACAAAGGCGCTTTTAAACAGTGTGTGAGAAATTAAGGTTTAC	636
Db	327	AAAGTGCAATGTGAAATTTTCGGGAAACACCTTTCTCAAATGGGAGAAAGCTGAGGCGCTCAC	386
QY	637	TCCTACTCTCACAGGAGAGAGCGCTTACAGTGCATACACAGACTGTCACACAGGCC	696
Db	387	AGCCTCCCGCAACACAGAGAGAGACCATATAGCTGCGCTCAGCTGCACCTGTGGCAGGTT	446
QY	697	TTTGTGTTCTTAAGTACAAATTTACAAAGGACATGGGTACTCACTTCTCTGAGAAAAACCCAC	756
Db	447	TTTGCTTTCCAAATACAGCTGTATAGGCATATGGCCACCATCTCAGCCGAGAAACCCAC	506
QY	757	AAAGTGAATTTATTTGTGAGAAATTTTACCGGAAAGATCATCTGAAAGAAATCACCTCCAT	816
Db	507	CAGTGTATGTACTGTGATTAAGATGTTTACCGCAAGGACCATCTGCGGAAACCATCTGCAG	566
QY	817	ACACAGCACCTTAAACAGAGAGCTTTAAGTGGGAAGATGTGGCAAGACTACATAATACC	876
Db	567	ACCCATGATCTTAAACAAAGAGGCGCTTCCACTGCTCTGAGTGCAGTAAGAAATTAACAATACG	626
QY	877	AAAGTGTGATTTTAAAGCTCACCTTGGCTTGCATGCGCAACAAAGTGGTGCACCTCACCTGT	936
Db	627	AAAGTGGGTACCGGCGCCACCTGGCCATGCACTGTCACAGACGGTGCACCTCAGCTGC	686
QY	937	AAAGTATGTTTGCACACTTTTGAAGACACGGAGTGTCTCTGAGACACCTTAAATCTCAT	996
Db	687	AAAGTGTGCTGCAGACCTTTGAGAGTACCAGGCGCTGCTAGAGCACCTGAAGGCCAC	746
QY	997	GAAGGCAAGTGTCTGTGGGTGTTAAAGAAAAAGCAACAGTGCAGCAATTTGTGATCGC	1056
Db	747	TCACGCGCGGTAGCAGCGCGTGCACAGGAGAGAGACACCCCTGTGACCACTCGCAGCCG	806
QY	1057	CGGTTCTACACCCGAAAGATGTCCGGAGACACATGCTGCTGCACACTGCGAAGAAAGGAC	1116
Db	807	CGGTTCTACTACTCTGTAGAGATGTACGCGCGCACTAGTGTGACACAGCGCGTAAGAC	866
QY	1117	TTCTCTGTCTGATTTGTGCACAGAGATTTGGGCGAAAGATCACTGCTGCACATATG	1176
Db	867	TTCTCTGTGCGAGTACTGTGCCAGCGGTTTGGCGGTAAAGACACCACTGACGCTCATGTC	926
QY	1177	AGAAGAGTCACAATCAAGAGCTTCTGAAGGTCAAAACAGAACAGTGTGATTTCTCTGAC	1236
Db	927	AGAAGAGGCACTCGCAGAGAGCTGTCTCAAGATCAAGACAGAGCGCGGTGGACATGTTAGGC	986
QY	1237	CCATTTTACCTGCAATGTGTGTGCTTATAAAGACGAGCTCTCTCCGCTGATGTCTTAA	1296
Db	987	CTACTCAGCTGAGCTCCACAGTGTGAGAGAGAGCTGAGCGCTGTGCTGTGATG	1046
QY	1297	CCTTCCAGTGA---ACTGTTATCAAAAGCATTACAAAACATTTTGTGAGTTAAACCTCTAC	1353
Db	1047	GCCTCTCGGACGTAATATGGGACCAAGGCTTCCCTGGCATGTTGCCCCATGGCATGTAT	1106
QY	1354	AACACTCCATTTTCACTGTCATGAGAGCTGGGATCTGCCCA---CCAAATGATCACAACT	1410
Db	1107	GGTGCCCAATCTCTTACCATGCCAGACGCGGCGTGCACACTCTCTTGGTGCACACAG	1166
QY	1411	TTACCTTTTGGGAATGATGCGCCCAATAGATATGGAACACTGTTTCTATCCCTCTCACACCTT	1470
Db	1167	CTGCCCATGGGTATGAGTACCTCTGGAATCCTCACTTATCTTCTTCCAGCTCAGCTC	1226
QY	1471	TCCTTCAATATCCGTTCTAGTTCTACTCATATGCAATTTTCTATTTCTGTAAGAAACAG	1530
Db	1227	CCTTCCAAATACAGCTTGGATCTACTCTCATCTTG-----CCGACAAA	1271
QY	1531	CCATTAAAGGGGAAATTTGAGAGTTACCTGAGTTTACAGGTGCGGTGCGCTCTTCTCA	1590



```
Db 1272 TTGCCAAAGTGGAGGTGGATAGTTTTCTGCGGGAGCTTCTCGAAGCCTCTCTCTCA 1331
Qy 1591 TCCCAAGATTCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCTCTCAGATTGGGTCC 1650
Db 1332 TCCGCTGAACCCAGCCGCTCA-----CCTCAGCCGGGGCA 1370
Qy 1651 CTAGATGATGGTCAGAGACCTCTCCCTATCCAAAAGCTCTATCTCCATCAGTGACCCC 1710
Db 1371 GCTCGGCCCTCTCTAGATGAAGCACTGCTTCCCAAGAGCCCGCCAACTCTCTGAGGCC 1430
Qy 1711 CTAACACACAGCATTGGATTTTCTCAGTTGTTTAAATTCATACCTTTAAATGGTCT 1770
Db 1431 CTCTGCGCTGTAATGGACTCTCCCACTCTGGGCTTTCTTCCACTCAACCTGCC 1490
Qy 1771 CCTATTA-----TCTCTATCAGTGGGAGCCTTGGAAATGAGCTATCCAGGAGAAGCA 1827
Db 1491 CCGTGAACCCACCTGGGGCCACAGGAGGCTGTCTATGGCTACTCCAGGCTGAGGCA 1550
Qy 1828 CATTCTTCTGTTCCAGCTCCCAACACACAGAGATCTTCAGGATCTCTGAAAACACT 1887
Db 1551 CAGCCCTGCTTACCCTTTGCAAGCTCAGCCTCAAGATTCCCAAGGAGCTGGGGACCA 1610
Qy 1888 ATAGGGCTTGGGTCTCTGCACTCACTGTGAGCAGCTTTCCAGCAGTAAAGCAAGT 1947
Db 1611 CTGAACCTTGGGCTCTGCACTCTTGGCTCTCTCTTCACTGCTGCGCTGAGT---AGC 1667
Qy 1948 ACCACCTCCCAAGTTTCCATCAAGCTTTTCAAGTAG 1983
Db 1668 ACCACCTGCTGCTTCCATCAAGCATTCCAGTAG 1703
```

## RESULT 14

## AK026936

## LOCUS

DEFINITION Homo sapiens cDNA: FLJ23283 fis. clone HEP08729, highly similar to AF006005 Homo sapiens zinc finger protein PLAGL2 mRNA.

## ACCESSION

## AK026936

## VERSION

## AK026936.1

## GI:10439912

## oligo capping; fis (full insert sequence).

## SOURCE

## Homo sapiens

## ORGANISM

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## 1 (sites)

## Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,

## Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,

## Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.

## NEDO human cDNA sequencing project

## Unpublished

## 2 (bases 1 to 2429)

## Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,

## Shibahara, T., Tanaka, T. and Nakamura, Y.

## Direct Submission

## Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,

## University of Tokyo, Laboratory of Genome Structure Analysis, Human

## Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,

## Japan (E-mail: cdnal@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,

## Fax: 81-3-5449-5416)

## NEDO human cDNA sequencing project supported by Ministry of

## International Trade and Industry of Japan; cDNA full insert

## sequencing. Research Association for Biotechnology; cDNA library

## construction, 5'- &amp; 3'-end one pass sequencing; Department of

## Virology and Human Genome Center, Institute of Medical Science,

## University of Tokyo (partly supported by Science and Technology

## Agency).

## Location/Qualifiers

## 1..2429

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /clone="HEP08729"

## /cell\_line="HepG2"

## /cell\_type="hepatoma"

## FEATURES

## source

/clone\_lib="HEP"

/note="Cloning vector pME18SFL3"

1..2429

/note="highly similar to AF006005 Homo sapiens zinc finger protein PLAGL2 mRNA"

BASE COUNT 540 a 743 c 599 g 547 t

## ORIGIN

## Query Match

Best Local Similarity 58.3%; Score 392.8; DB 9; Length 2429;

Matches 826; Conservative 0; Mismatches 542; Indels 48; Gaps 6;

```
Qy 577 AACTTTCTTCCAACTGTGTGACAGGCCCTTAAAGTGTACAAAGAGCTGAGAAATTAAGGTTTCA 636
Db 199 AAGTGCCAAATGTGAATTTTCGGGAACACCTTCTCAAAATGGGAGAGCTGAGGCTTCA 258
Qy 637 TCCTACTCTCACAGGAGAGAGGCCCTTAAAGTGTACAAAGAGCTGAGAAATTAAGGTTTCA 696
Db 259 AGCTTCCGCAACCCAGAGCAGACCATATAGTGCCTCAGCTGACACTGAGGCAAGGCT 318
Qy 697 TTTGTTTCTTAAGTACAAATTAAGGACATGGTACTACTCATTTCTCTGAGAAATTAAGGTTTCA 756
Db 319 TTTGCTTCCAAATTAAGTGTATAGGCACATGCCACCTCAGCTGAGGCAAGGCTTCA 378
Qy 757 AAGTGAATTTATTTGTCAGAAATTTTACCCGAAAGATCATCTGAAGATCATCTTCAAT 816
Db 379 CAGTGATGTATGTATAGATTTTACCCGAAAGATCATCTGAGGCAAGGCTTCA 438
Qy 817 ACACAGACCTTAAACAAAGAGAGCTTTAAAGTGTGGAAGATGTGCGAAGACTTACAATACC 876
Db 439 ACCATGATCTTAAACAAAGAGGCTTCACTGCTCTGAGTGGTGAAGATTAACAATACG 498
Qy 877 AAGCTTGGATTTAAAGCTCACTTGGCTTGCATGCGCAACAAAGTGGTGAAGTCACTTCACTGT 936
Db 499 AAGCTGGGTACCGGCGCCACCTGGCCATGCTGCTGCCAGAGCGGTGACCTCAGCTGC 558
Qy 937 AAGGTATGTTTGCMAACTTTTGAAGACGCGAGTGTCTTCTGGAGCACCTTAAATCTCAT 996
Db 559 AAGGTGTGCTGCAGACCTTTGAGAGTACCCAGGCGCTGCTAGAGCACCTGAAGGCCAC 618
Qy 997 GCAGGCAAGTCTCTGTGGGTTTAAAGAAAAAGACCACTGCGCAACATTTGATTCG 1056
Db 619 TCACGCGGTAGCAGGCGGTCCAGGAGAGAGACCCCTGTGACCACTGCGACCG 678
Qy 1057 CGTTTCAACCCGAAAGAGTCCGCGACACATGTTGGTGTGCACACTGGAAGAGAGAC 1116
Db 679 CGTTTCTATCTGTAAGATGTACGCGGACCTAGTGTGTCACACAGCGCGTGAAGAC 738
Qy 1117 TTCCTCTGTCAGTATTGTGCACAGAGATTGGGCGAAAGGATCACTGACTCGACATATG 1176
Db 739 TTCCTGTGCCAGTACTGTGCCCTAGCGGTTTGGCCGTAAGGACCACTGACGCTCATGTC 798
Qy 1177 AAGAGAGTCAAAATCAAGAGCTTCTGAAGTCAAAACAGAGACAGTGGATTTCCCTTAC 1236
Db 799 AAGAGAGCCTACTCGCAGAGCTGCTCAAGATCAAGACAGAGCGCGTGAAGATGTTAGC 858
Qy 1237 CCATTACTGCGAATGTCTGTGCTATAAAGACGAGCTCTTCCCGGTGATGCTCTTA 1296
Db 859 CTACTCAGCTGAGCTCCACAGTCAAGTGTGAAGGAGAGCTGAGCCCTGTGCTGCAATG 918
Qy 1297 CTTTCCAGTGA---ACTGTTATCAAGCATTCAAAACACTTTTGAAGTAAACCTTAC 1353
Db 919 GCCTCTCGGAGCGTAATGGGAGCAAGGCTTCCCTGTCATGTTGCCCATGGCATGTAT 978
Qy 1354 AACACTCCATTTCAGTCCATGACAGCTCGGATCTGCCCA---CCAATGATCAACT 1410
Db 979 GGTGCCACATCCCTACATGCCAGCAGCGGCGTGCACACTCCCTGGTGCACAACAG 1038
Qy 1411 TTACTTTTGGGAATGACATGCCCAATAGATATGACACTGTTTCACTCCCTCTCACACCTT 1470
Db 1039 CTGCCCATGGTATGAGTACCTCTGGAATCCTCAGCTATCTTCTCCCGAGCTCAGCTC 1098
Qy 1471 TCCTTCAATATCGGTTCAAGTTCTTCACTCATATGCAATTTCTATTCTCTGAAAAAGACAG 1530
```

```

Db      1099  CCTCAAAATACAGCTTGAGTCTACCTCACTCTG-----CCGACAAA 1143
QY      1531  CCATTAAAGGGGAAATGAGAGTTTACCTGATGGAGTTACAAGTGGCGTGCCTCTTCA 1590
Db      1144  TTGCCCAAGTGGAGTGGATAGTTTCTGGCGGAGCTTCTCGGAAGCCTGTCCCTCTCA 1203
QY      1591  TCCCAAGATTTCAAGCATCGTCATCATCTAAAGCTAGGTTGGATCTCAGATTGGGTCC 1650
Db      1204  TCCGCTGAACCCAGCCGCTCA-----CCTCAGCCGGCGGCA 1242
QY      1651  CTAGATGATGGTGCAGGACCTCTCCCTATCCAAAGCTCTATCTCCATCAGTGACCCC 1710
Db      1243  GTTGGGCCCCCTCTAGTGAAGCACTGCTTGCAGAGCCCGCCCAACCTCTCTAGGCC 1302
QY      1711  CTAACACACACAGCAATGGATTTTCTCAGTTGTTTAAATTTTCATACCTTTAAATGTCCT 1770
Db      1303  CTCTGGCTGCTAAATGTGACCTTCTCCACCTACTGGGCTTTCTTCCACTCAACCTGCC 1362
QY      1771  CCTATATA---TCCTCTATCAGTGGGAGCTTGGATGAGCTATTCCAGGAAGACA 1827
Db      1363  CGGTGTAACCCACTGGGGCCACAGAGGCTGTCTCATGGGCTACTCCAGGCTGAGGCA 1422
QY      1828  CATCTCTCTGTTTCCAGCTCCCAACACACAGATCTTCCAGGATCTCTGCAAAACACT 1887
Db      1423  CAGCCCTGCTTACCCTTTGCAAGCTCAGCTCAAGATTCCTCCAGAGCTGGGGACCA 1482
QY      1888  ATAGGGCTTGGGTCTCTGCACTCACTGTGACAGCTTTTCCAGGAGTTTAAGCAAGT 1947
Db      1483  CTGAACCTTTGGGCTCTGCACTCTTGGCTCTCTTCCAGCTGTGGCTGAGT---AGC 1539
QY      1948  ACCACCTTCCAGCTTTCATCAAGCTTTTCAGTAG 1983
Db      1540  ACCACCTGCTGCTTTCATCAAGCATTCAGTAG 1575

RESULT 15
LOCUS   AF006005 4051 bp mRNA linear PRI 07-MAR-2000
DEFINITION Homo sapiens zinc finger protein PLAGL2 (PLAGL2) mRNA, complete
cds.
ACCESSION AF006005
VERSION AF006005.1 GI:3513455
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4051)
AUTHORS Kas,K., Voz,M.L., Hensen,K., Meyen,E. and Van de Ven,W.J.
TITLE Transcriptional activation capacity of the novel PLAG family of
zinc finger proteins
JOURNAL J. Biol. Chem. 273 (36), 23026-23032 (1998)
MEDLINE 98389728
PUBMED 9722527
REFERENCE 2 (bases 1 to 4051)
AUTHORS Kas,K., Hensen,K., Meyen,E., Voz,M.L. and Van de Ven,W.J.M.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-1997) Laboratory for Molecular Oncology, Center
of Human Genetics, Herestraat 49, Leuven 3000, Belgium
FEATURES
source
1..4051
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q11"
1..4051
/gene="PLAGL2"
203..1693
/gene="PLAGL2"
/note="C2H2-type zinc finger protein; PLAG-like 2"
/codon_start=1

gene
CDS
1099 1143
1531 1590
1144 1203
1591 1650
1204 1242
1651 1710
1243 1302
1711 1770
1303 1362
1771 1827
1363 1422
1828 1887
1423 1482
1888 1947
1483 1539
1948 1983
1540 1575
```

```

/product="zinc finger protein PLAGL2"
/protein_id="AAC34252.1"
/db_xref="GI:3513456"
/translation="MTTFTSVPMWIDAKOEVEVWKLVPKPRGREAESQVKQCEI
/TPSPNGKLRPHSLPQEPORPSCPOLHCGKAFASKVLYRHATHSAOKPHOCMY
CDKPHRRKHLRNLQTHDPNKAULHSCCKKNYNTKLYRRHLAMHAAASGDLCKV
CLQPTFESTQALLHLKHAHSRRVAGAKKKPKDCHRRFYTRKRRHLVHTGRKD
FLCOYCAORFGRKHLRHVKSHQELLKIKTEFVMDLGLLSCSTVSVBELSPVL
CMASRDVMTKAFPGMLPMWYGAHPTMSTGVPKSHLHNTLPMGMSYPLESSPIS
EAQLPKYQLGSTSYLPDKLPKVEVDFSLAELPGLSLSSAPQASPOPAAAAALLD
EAALLKSPANLSEALCAANVDFSHLGLPLNLPKCPNPGATGHLVMGYSQBAQPLL
TTLQAPQDPSGAGPLNFPPLHSLPFTSTLSSTLTPRFHQAFQ"
BASE COUNT 874 a 1192 c 972 g 1013 t
ORIGIN
Query Match 5.4%; Score 392.8; DB 9; Length 4051;
Best Local Similarity 58.3%; Pred. No. 6.8e-60;
Matches 826; Conservative 0; Mismatches 542; Indels 48; Gaps 6;
QY 577 AACTTTCCTTGGCAACTGTGTGACAAGGCTTTTAAAGTGTGTGAGAAAATTAAAGTTTCAC 636
Db 317 AAGTGCATGTGAATTTTCGGGAACACCTTTCTCAATGGGAGAGAGCTGAGGCTTCAC 376
QY 637 TCTTACTCTCACACAGAGAGAGGCTTCAAGTGCATACAAAGACTGCACCAAGGCC 696
Db 377 AGCCTCCGCAACACAGAGAGAGACCATATAGCTGCCCTCAGCTGCATCTGTGCAAGGCT 436
QY 697 TTTGTTTCTTAAGTACAAATTAACAAGGCACATGGCTACTCATTTCTCTCGTGAAGAAACCCAC 756
Db 437 TTTGCTTCCAAATACAAAGCTGTATAGGCACTGGCCACCCACTCAGCCCAAGAACCCAC 496
QY 757 AAGTGTAAATTAATTTGTGAGAAAATGTTTACCGGAAAGATCATCTGAAGATCACTCCAT 816
Db 497 CAGTGTATGATCTGTGTAAGATGTTTACCGCAAGGACCATCTGCGGAACCATCTGCAG 556
QY 817 ACACAGACCTTAAACAAAGAGAGCTTTAAGTGCAGAGATGTGGCAGACTACATACC 876
Db 557 ACCCATGATCTTAAACAAAGAGGCTTCCACTGTCTGAGTGGGTAAAGATTAACAATACG 616
QY 877 AAGCTTGGATTTAAAGCTCACTTGGCTTGCATCGCAACAAAGTGGTGAACCTCACCTGT 936
Db 617 AAGCTGGGCTACCGGGCCACCTGGCCCATGCTGCTGAGAGTACCCAGGCTGTGACGCTGC 676
QY 937 AAGTATGTTTGCAAACTTTTGAAGCAGCGGAGTGTCTTGGAGCAGCTTAAATCTCAT 996
Db 677 AAGTGTGCTGTGAGACCTTTGAGAGTACCCAGGCTGTCTAGAGCACTTGAAGGCCAC 736
QY 997 GCAGGCAAGTCTGTGTGGGGTTTAAAGAAAAGAACACAGTGCAGCACTTGGATGCGC 1056
Db 737 TCACGCGGGTAGCAGCGGCTGCCAAGAGAAAGACACCCCTGTGACCACTTGGACCGG 796
QY 1057 CGGTTCTACACCCGAAAGAGTGTCCGAGACACATGCTGGTGTGCACACTTGAAGAAAGAC 1116
Db 797 CGGTTCTATCTCGTAAGGATGTACGGCGGACCTAGTGGTGCACACAGGCCGTGAAGAC 856
QY 1117 TTCCTCTGTCTCAGTATTTGTGCACAGAGATTTGGGCGAAAGGATCACTGACCTGCACATG 1176
Db 857 TTCCTGTGCCAGTACTGTGCCAGCGTGTGGCGGTAAAGGACCACTGACGCGTCATGTC 916
QY 1177 AAGAGAGTCTCAATCAAGACTTCTGAAGTCAAAACAGACAGGAGTGTTCCTTTCAG 1236
Db 917 AAGAAGAGCCACTCGCAGGAGCTCTCAAGATCAAGACAGAGCCCGGTGGACATGTTAGGC 976
QY 1237 CCATTTTACCTGCAATGTGTCTGTGCTTATAAAGACGAGCTTCTTCGGTGTGATCTCTTA 1296
Db 977 CTACTCAGCTGCAGCTCCACAGTCACTGTGAGGAGAGGCTGAGCCCTGTCTGTGTCATG 1036
QY 1297 CCTTCCAGTGA---ACTGTTATCAAGCCATTCTCAAAACCTTTGAGTTAAACCTCTAC 1353
Db 1037 GCCTCTCGGGACGTAAATGGGAGCAAGGCTTCCCTGGCATGTTGGCCATGGCAGTAT 1096
QY 1354 AACACTCATTTCAGTCCATGCAGAGCTCGGGATCTGCCCA---CCAAATGATCACAACT 1410
```

Db 1097 GGTGCCACATCCCTACCATGCCAGCACGGCGTGCCACACTCCCTGGTGCAACACG 1156  
Qy 1411 TTACCTTTGGGAATGACATGCCCAATAGATATGGACATGTTTCATCCCTCTCACACCTT 1470  
Db 1157 CTGCCATGGGATGAGCTACCCCTCTGGAATCCTCACCTATCTCTCCCGAGCTCAGCTC 1216  
Qy 1471 TCTTTCAAAATATCCGTTTCTAGTTTCTACCTCATATGCAATTTCTATTCTGAAAGAAAGACAG 1530  
Db 1217 CCTCCAAAATACCACTGGATCTACCTCATACTTG-----CCCGACAAA 1261  
Qy 1531 CCAITTAAGGGGGAATTCAGAGTTTACCTGATGGAGTTACAAGGTGGCGTGCCCTCTTCA 1590  
Db 1262 TTGCCCAAGTGGAGTGGATGTTTCTCTGGCGGAGCTTCTGGAAGCTGTCTCTCTCA 1321  
Qy 1591 TCCCAAGATTCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCTCAGATTGGGTCC 1650  
Db 1322 TCCGTGTAACCCAGCCCGCCTCA-----CCTCAGCCGGCGCA 1360  
Qy 1651 CTAGATGATGTGCGAGGAGCTCTCCCTATCCAAAAGCTTATCTCCATCAGTGACCCC 1710  
Db 1361 GCTGGGCCCTCCTAGATGAGCACTGTGTGCCAAGAGCCCGCCAACTCTCTGAGGCC 1420  
Qy 1711 CTAAACACACAGCATTTGGATTTTCTCAGTTGTTTAAATTTCATACCTTTAAAGTCTCT 1770  
Db 1421 CTCTGCGCTGCTAATGTGGACTTCTCCACCTACTTGGGCTTCTTCCACTCAACCTGCC 1480  
Qy 1771 CCCTATTA--TCCTCTATCAGTGGGAGCCCTTGAATGAGCTATTCGCCGGAAGACA 1827  
Db 1481 CCGTGTAAACCCACCTGGGGCCACAGAGGCTGGTCTATGGGCTACTCCGAGCTGAGCA 1540  
Qy 1828 CATTTCTCTGTTTCCAGCTCCCCACAAACACAGGATCTTCAGGATCCTGCAAACT 1887  
Db 1541 CAGCCCTGCTTACCACTTTGCAAGCTCAGCCTCAAGATTCCCCAGAGCTGGGGGCA 1600  
Qy 1888 ATAGGCTTGGGTCTCTGCACTCACTGTCAAGAGCTTTCAAGCAGTTTAAAGCACAAGT 1947  
Db 1601 CTGAACCTTGGGCTCTGCACTCTCTGCTCTGCTCTTCAAGCTCTGGCTGAGT---AGC 1657  
Qy 1948 ACCACCTCCAGCTTCCATCAGCTTTTTCAGTAG 1983  
Db 1658 ACCACCTGCTGCTTCCATCAGCATTCAGTAG 1693

Search completed: January 31, 2004, 13:36:46  
Job time : 25657 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 03:59:43 ; Search time 1682 Seconds  
(without alignments)  
11736.617 Million cell updates/sec

Title: US-09-242-772-116  
Perfect score: 7313  
Sequence: 1 ggcagcgatcacatacaat.....tatgaataaaatctgtgac 7313

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq 13Jun03.\*

```

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7312.6	100.0	7313	19	AAV29268
2	351.6	4.8	2334	19	AAV18481
3	351.6	4.8	2828	24	ABN86525
4	350	4.8	2815	24	ABL61932
5	350	4.8	4632	24	ABN86524
6	301.8	4.1	5028	24	ABN86508
7	294.4	4.0	2790	19	AAV18480
8	294.4	4.0	3975	24	ABN86526

9	277.6	3.8	2561	19	AAV29269	Nucleotide sequenc
10	277.6	3.8	2561	24	ABT11028	Human breast cance
11	276	3.8	2738	20	AAZ33549	Human breast tumou
c 12	267	3.7	475	22	ABA52644	Human foetal liver
c 13	267	3.7	475	22	ABA22432	Probe #898 for gen
c 14	267	3.7	475	22	AAK00905	Human brain expres
c 15	267	3.7	475	22	AAK26360	Human bone marrow
c 16	267	3.7	475	22	AAI10993	Human bone marrow
c 17	267	3.7	475	22	AAI32253	Probe #926 for gen
c 18	267	3.7	475	22	AAI00914	Probe #905 used to
c 19	267	3.7	475	23	ABS25951	Human liver single
c 20	267	3.7	475	24	ABS00948	Human genome-deriv
c 21	131.4	1.8	230	22	ABA65256	Human foetal liver
c 22	131.4	1.8	230	22	ABA32357	Probe #10823 for g
c 23	131.4	1.8	230	22	AAK13673	Human brain expres
c 24	131.4	1.8	230	22	AAK39414	Human bone marrow
c 25	131.4	1.8	230	22	AAI20227	Probe #10160 for g
c 26	131.4	1.8	230	22	AAI45427	Probe #14113 used
c 27	131.4	1.8	230	22	AAI05931	Probe #5922 used t
c 28	131.4	1.8	230	23	ABS39002	Human liver single
c 29	131.4	1.8	230	24	ABS13501	Human genome-deriv
30	126.8	1.7	324	24	ABL77615	Human ovarian canc
31	74.6	1.0	8056	25	ABZ10246	Haematopoietic cel
32	72.4	1.0	12237	24	ABL34358	Human immune syste
33	72	1.0	72	22	AAI13082	DNA encoding zinc
34	72	1.0	72	25	ACC41633	Human zinc finger
c 35	71	1.0	8056	25	ABZ10246	Haematopoietic cel
c 36	65.8	0.9	2864	22	AAH14677	Human cDNA sequenc
c 37	65.4	0.9	7348	22	AAZ46336	Tumour suppressor
c 38	64.8	0.9	8056	25	ABZ10100	Haematopoietic cel
c 39	63.6	0.9	113515	24	ABL34175	Human immune syste
40	62.8	0.9	9504	24	ABK28407	DNA transcription
41	62.4	0.9	703	22	AAH16178	Human cDNA clone (
42	62.4	0.9	2729	22	AAH06346	Human cDNA sequenc
43	62.4	0.9	5893	24	ABL32859	Human immune syste
44	61.6	0.8	6301	24	ABN80100	Human chemically m
45	61.4	0.8	5544	24	ABL70478	Chemically treated

#### ALIGNMENTS

RESULT 1  
AAV29268  
ID AAV29268 standard; cDNA; 7313 BP.  
XX  
AC AAV29268;  
XX  
DT 21-AUG-1998 (first entry)  
XX  
DE Nucleotide sequence of human PLAG1.  
XX  
Human PLAG1 gene; PLAG1; tumorigenesis gene; T-gene; PLAG2; CTNNB1;  
KW antibody; benign tumour; malignant tumour; leukaemia; lymphoma;  
KW cancer; inhibition; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT CDS  
FT Location/Qualifiers  
FT 481..1983  
FT /\*tag= a  
FT /product= "PLAG1 protein"  
FT /\*tag= b  
FT /transl\_except= (pos: 1603..1605, aa: Gly)  
FT /\*tag= c  
FT /transl\_except= (pos: 1861..1863, aa: Gly)

97EP-0200130.

PR 22-AUG-1996; 96EP-0202339.  
XX (KULE-) KU LEUVEN RES & DEV.  
PA (UYGO-) UNIV GOETTERBORGS HOLDINGBOLAGET AB.  
PA (LEUV-) LEUVEN RES & DEV.  
XX  
PI Kas KP, Stenman KGD, Van De Ven WJM, Voz ML;  
XX P-PSDB; AAM37948.  
XX WPI; 1998-132252/13.  
XX P-PSDB; AAM37948.  
XX  
XX New tumorigenesis T-genes and proteins - useful for, e.g. preparing  
PT antibodies for clinically diagnosing cells having non-physiological  
PT proliferative capacity such as lipoblastomas  
XX  
XX Claim 2; Fig 4; 71pp; English.  
XX  
XX This is the nucleotide sequence of the human PLAG1 gene. It is a  
CC tumorigenesis gene (T-gene), which is isolated in the form of PLAG1,  
CC PLAG2, and CTNNB1 genes. Their proteins can be used as a starting  
CC point for preparing antibodies for clinically/medically diagnosing  
CC cells having a non-physiological proliferative capacity as compared  
CC to wild type cells, where the former cells are selected from both  
CC benign and malignant tumours, as well as leukaemia and lymphomas.  
CC Derivatives of the T-gene are also used in the diagnosis and  
CC preparation of therapeutical compositions for the treatment of cancers,  
CC such as nucleic acid derivatives, and antibodies. The T-gene may be  
CC used as a starting point for designing suitable expression-modulating  
CC compounds or techniques for the treatment of non-physiological  
CC proliferation phenomena in humans or animals. Expression inhibitors  
CC of the T-gene can be used in the treatment of diseases involving  
CC benign or malignant tumours.  
XX  
SQ Sequence 7313 BP; 2287 A; 1327 C; 1356 G; 2342 T; 1 other;

Query Match 100.0%; Score 7312.6; DB 19; Length 7313;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 7312; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGCAGCGCATACACTACAAATGGCTGCTGGAAGAGCGGTAAAGAAACAATTCAGGCC 60  
DB 1 GGCAGCGCATACACTACAAATGGCTGCTGGAAGAGCGGTAAAGAAACAATTCAGGCC 60  
QY 61 GCGCGTCCAGCCGCAATATGAGAAAAATTTAGAAATTCGCGGCGGTGTAGAG 120  
DB 61 GCGCGTCCAGCCGCAATATGAGAAAAATTTAGAAATTCGCGGCGGTGTAGAG 120  
QY 121 GCGCGGACGGCGGAGGAGGATGTAAAGCCCGCGGTGCTCTTGGTGTGCTT 180  
DB 121 GCGCGGACGGCGGAGGAGGATGTAAAGCCCGCGGTGCTCTTGGTGTGCTT 180  
QY 181 GCGCGTATTTGGCAACCAGAAATGCTTCAATCTGTGACGGTCTAATAAGTTGCTT 240  
DB 181 GCGCGTATTTGGCAACCAGAAATGCTTCAATCTGTGACGGTCTAATAAGTTGCTT 240  
QY 241 CTAGAGTTTGGACAGGCGCTCAGATTGGCCAAATCGGAGGATTCGATTCACACTCT 300  
DB 241 CTAGAGTTTGGACAGGCGCTCAGATTGGCCAAATCGGAGGATTCGATTCACACTCT 300  
QY 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAAGTCTGAGGCTTTTCCATCAGTAAT 360  
DB 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAAGTCTGAGGCTTTTCCATCAGTAAT 360  
QY 361 CAGTCCCTTTTGTCTTTTACGACCAATGAAATCTGAGAGGCGCACTTAAGCTATA 420  
DB 361 CAGTCCCTTTTGTCTTTTACGACCAATGAAATCTGAGAGGCGCACTTAAGCTATA 420  
QY 421 TCATTTAGTGGAGTTGGCAGTCCAGTGTCCAAACAGAGCGCTGTTTAGCTGCG 480  
DB 421 TCATTTAGTGGAGTTGGCAGTCCAGTGTCCAAACAGAGCGCTGTTTAGCTGCG 480  
QY 481 ATGGCCACTGTCAATTCCTGGTGAATTTGTGAGAGTAAAGATACCCAGAAAGTCCCTTCA 540

DB 481 ATGGCCACTGTCAATTCCTGGTGAATTTGTGAGAGTAAAGATACCCAGAAAGTCCCTTCA 540  
QY 541 GGGAAACCTAAGCGTGGTGAACCAACCAAGAAAAAACTTTCTTGGCAACTGTGTGAC 600  
DB 541 GGGAAACCTAAGCGTGGTGAACCAACCAAGAAAAAACTTTCTTGGCAACTGTGTGAC 600  
QY 601 AAGCCCTTTAAACAGTGTGAGAAAATTAAGGTTCACTCTACTCTCACACAGGAGAGG 660  
DB 601 AAGCCCTTTAAACAGTGTGAGAAAATTAAGGTTCACTCTACTCTCACACAGGAGAGG 660  
QY 661 CCCTACAAAGTGCATACAAAGACTGCACCAAGGCTTTGTTTCTAAGTACAAATACAA 720  
DB 661 CCCTACAAAGTGCATACAAAGACTGCACCAAGGCTTTGTTTCTAAGTACAAATACAA 720  
QY 721 AGGCACATGGCTACTCAATTCCTCGAGAAAAACCCACAGTGTAAATTTGTGAGAAATG 780  
DB 721 AGGCACATGGCTACTCAATTCCTCGAGAAAAACCCACAGTGTAAATTTGTGAGAAATG 780  
QY 781 TTTCCACCGGAAAGATCACTCTGAAGAAATCACTCTCTACACAGACCTTAACAAAGAGC 840  
DB 781 TTTCCACCGGAAAGATCACTCTGAAGAAATCACTCTCTACACAGACCTTAACAAAGAGC 840  
QY 841 TTTAAGTGCAGAAATGTGGCAAGAACTACAAATCCAAAGCTTGGATTTAAACCTCACTG 900  
DB 841 TTTAAGTGCAGAAATGTGGCAAGAACTACAAATCCAAAGCTTGGATTTAAACCTCACTG 900  
QY 901 GCCTTGCATGCCGCAACAAAGTGTGACCTCACTGTAAAGTATGTTTGAACCTTTTGA 960  
DB 901 GCCTTGCATGCCGCAACAAAGTGTGACCTCACTGTAAAGTATGTTTGAACCTTTTGA 960  
QY 961 AGCAGCGGAGTGTCTTGTGAGCACTTAAATCTCATGAGGCAAGTCTGTGGTGGGTT 1020  
DB 961 AGCAGCGGAGTGTCTTGTGAGCACTTAAATCTCATGAGGCAAGTCTGTGGTGGGTT 1020  
QY 1021 AAGAAAAAAGACACAGTGCAGAACTGTGATCGCGGTTCTACACCCGAAAGAGTGC 1080  
DB 1021 AAGAAAAAAGACACAGTGCAGAACTGTGATCGCGGTTCTACACCCGAAAGAGTGC 1080  
QY 1081 CGGAGACACATGTGTGACACTGGAAGAAAGAGTCTCTCTGTGATTTGTGACAG 1140  
DB 1081 CGGAGACACATGTGTGACACTGGAAGAAAGAGTCTCTCTGTGATTTGTGACAG 1140  
QY 1141 AGATTTGGCGAAAGAGTCACTGACATATGAAGAAAGAGTCAATCAAGAGCTT 1200  
DB 1141 AGATTTGGCGAAAGAGTCACTGACATATGAAGAAAGAGTCAATCAAGAGCTT 1200  
QY 1201 CTGAAGTCAAAACAGAACAGTGGATTTCTTGGACCATTTACCTGCAATGTGTGTG 1260  
DB 1201 CTGAAGTCAAAACAGAACAGTGGATTTCTTGGACCATTTACCTGCAATGTGTGTG 1260  
QY 1261 CCTATAAAGACAGAGTCTCTTCCGGTGTGATGCTCTTACCTTCAGTGAACCTTATCAAAG 1320  
DB 1261 CCTATAAAGACAGAGTCTCTTCCGGTGTGATGCTCTTACCTTCAGTGAACCTTATCAAAG 1320  
QY 1321 CCATTCACAAACACTTTGAGTAAACCTCTAAGACTCACTCACTCACTCACTCACTCACT 1380  
DB 1321 CCATTCACAAACACTTTGAGTAAACCTCTAAGACTCACTCACTCACTCACTCACTCACT 1380  
QY 1381 TCGGGATCTGCCCAACCAATGATCAAACTTTTGGAAATGACATGCCCAATAGAT 1440  
DB 1381 TCGGGATCTGCCCAACCAATGATCAAACTTTTGGAAATGACATGCCCAATAGAT 1440  
QY 1441 ATGGACACTGTTCATCTCTCTCACCACTTTCTTCAAATATCCGTTCACTTCACTCA 1500  
DB 1441 ATGGACACTGTTCATCTCTCTCACCACTTTCTTCAAATATCCGTTCACTTCACTCA 1500  
QY 1501 TATGCAATTTCTATTCCTGAAAAAGACAGCATTTAAAGGGGAAATTTAGAGTTTACCTG 1560  
DB 1501 TATGCAATTTCTATTCCTGAAAAAGACAGCATTTAAAGGGGAAATTTAGAGTTTACCTG 1560  
QY 1561 ATGGAGTTACAGGTGGCGCTCTTCTCATCCCAAGATTTCTCAAGCATCTGTCACTCATCT 1620  
DB 1561 ATGGAGTTACAGGTGGCGCTCTTCTCATCCCAAGATTTCTCAAGCATCTGTCACTCATCT 1620

1621 AAGCTAGGGTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGGAGACCTCTCCCTA 1680  
1621 AAGCTAGGGTGGATCCTCAGATTGGGTCCCTAGATGATGGTGCAGGAGACCTCTCCCTA 1680  
1681 TCCAAAGAGCTCTATCTCCATCAGTACAGCCCTTAAACACACAGAGATTGGATTTTCTCAG 1740  
1681 TCCAAAGAGCTCTATCTCCATCAGTACAGCCCTTAAACACACAGAGATTGGATTTTCTCAG 1740  
1741 TTGTTTAAATTTTATACCTTTAAATGGTCCCTCTAATATCTCTATCAGTGGGAGCCTT 1800  
1741 TTGTTTAAATTTTATACCTTTAAATGGTCCCTCTAATATCTCTATCAGTGGGAGCCTT 1800  
1801 GGAATGAGCTATTCCAGGAGAGACACATTTCTTTCTGTTTCCAGCTCCCCACACAAACA 1860  
1801 GGAATGAGCTATTCCAGGAGAGACACATTTCTTTCTGTTTCCAGCTCCCCACACAAACA 1860  
1861 CAGGATCTTCAGGATCCTGCAACACTATAGGGCTTTGGGTCTCTGCACTCACTGTGACGA 1920  
1861 CAGGATCTTCAGGATCCTGCAACACTATAGGGCTTTGGGTCTCTGCACTCACTGTGACGA 1920  
1921 GCTTTCCAGCAGAGTTTAAAGCAAGTACACACCTCCAGCTTTCCATCAAGCTTTTCTAG 1980  
1921 GCTTTCCAGCAGAGTTTAAAGCAAGTACACACCTCCAGCTTTCCATCAAGCTTTTCTAG 1980  
1981 TAGGATCTCGGACATGGATTCATTACAGAAATGTATGTAGCTGTGCCCTAGATGACC 2040  
1981 TAGGATCTCGGACATGGATTCATTACAGAAATGTATGTAGCTGTGCCCTAGATGACC 2040  
2041 ATTATTTATTTAGTGGCTACTTTAAACAGATATAAAAAATTTCTGCTTTTGTATATACAA 2100  
2041 ATTATTTATTTAGTGGCTACTTTAAACAGATATAAAAAATTTCTGCTTTTGTATATACAA 2100  
2101 ATTTTCATTAAAGCCAGTATAAAATAGAACTAGCTTTTAAATCTGAGCTTTGGAACATTT 2160  
2101 ATTTTCATTAAAGCCAGTATAAAATAGAACTAGCTTTTAAATCTGAGCTTTGGAACATTT 2160  
2161 GTGTTCAAGTTAAGTTTACCTGGGTATTTTGCTGATTCAGTCCCAATGTGCACATTTTA 2220  
2161 GTGTTCAAGTTAAGTTTACCTGGGTATTTTGCTGATTCAGTCCCAATGTGCACATTTTA 2220  
2221 AGACTTTTATTTTCCATATAGGAAGCAATATTAGTAGTAAATCTTTTACAAATCCCA 2280  
2221 AGACTTTTATTTTCCATATAGGAAGCAATATTAGTAGTAAATCTTTTACAAATCCCA 2280  
2281 TTTTCAAAATTAATTTAGATCTTAAATTTTCAATTTTGTCTAATTAACAGTGGCTCTACC 2340  
2281 TTTTCAAAATTAATTTAGATCTTAAATTTTCAATTTTGTCTAATTAACAGTGGCTCTACC 2340  
2341 TTTTGACATCTGGCTCAATTAATAAATTTAGCAATAGAAATGTAAATTTGTATAAAAAGTTTG 2400  
2341 TTTTGACATCTGGCTCAATTAATAAATTTAGCAATAGAAATGTAAATTTGTATAAAAAGTTTG 2400  
2401 TGAATAACTCAAGGTTTAAATTTTCTTACTAGCTTCTAAATGGAATTAATCAAGTGC 2460  
2401 TGAATAACTCAAGGTTTAAATTTTCTTACTAGCTTCTAAATGGAATTAATCAAGTGC 2460  
2461 TTTCAAAATTAATTTAGATCTTAAATTTTCAATTTTGTCTAATTAACAGTGGCTCTACC 2520  
2461 TTTCAAAATTAATTTAGATCTTAAATTTTCAATTTTGTCTAATTAACAGTGGCTCTACC 2520  
2521 TTTTGACATCTGGCTCAATTAATAAATTTAGCAATAGAAATGTAAATTTGTATAAAAAGTTTG 2580  
2521 TTTTGACATCTGGCTCAATTAATAAATTTAGCAATAGAAATGTAAATTTGTATAAAAAGTTTG 2580  
2581 TTAGGACCTTTAGGAAAACAAACCTTAAACACAAAGCATCAGTATTTATAGCAAAAG 2640  
2581 TTAGGACCTTTAGGAAAACAAACCTTAAACACAAAGCATCAGTATTTATAGCAAAAG 2640  
2641 AGACTCTGTTAGGTGACATGGCATTTTCTGTCACCTTAATAGTGTGGCCCTTAAATTTAGTACA 2700  
2641 AGACTCTGTTAGGTGACATGGCATTTTCTGTCACCTTAATAGTGTGGCCCTTAAATTTAGTACA 2700

2701 CAGGATATTTTGTGCGTGTTCATCTCTTTAAACATGCTATCTTTTCAATTTAATAATAGTA 2760  
2701 CAGGATATTTTGTGCGTGTTCATCTCTTTAAACATGCTATCTTTTCAATTTAATAATAGTA 2760  
2761 ATAGTGTATGGCATTTGGGTCCTTTCAGAGTCGATATATAGGTAGATCTCTTTAGTCTTTTC 2820  
2761 ATAGTGTATGGCATTTGGGTCCTTTCAGAGTCGATATATAGGTAGATCTCTTTAGTCTTTTC 2820  
2821 CACCTTTTACATCCAAAGGGTGGTCAAGTCAGCCAGCAATTTATTTTCAATTTGTGGCC 2880  
2821 CACCTTTTACATCCAAAGGGTGGTCAAGTCAGCCAGCAATTTATTTTCAITTTGTGGCC 2880  
2881 CAGGTTAGTCCATATCTTAGGCCATTTGTGAACTGTGCAAGTCAGGTGTGTATTATCCC 2940  
2881 CAGGTTAGTCCATATCTTAGGCCATTTGTGAACTGTGCAAGTCAGGTGTGTATTATCCC 2940  
2941 ACAGTGGATTGACTCAGCCTCTGTGGTGCAGACCTTTCTTAAGCAGGAAGATAGACGTGAA 3000  
2941 ACAGTGGATTGACTCAGCCTCTGTGGTGCAGACCTTTCTTAAGCAGGAAGATAGACGTGAA 3000  
3001 GCACATGGTTACATTTGGGAACTTTGTAGGATCATGGCCCTGTAGCCAGGTTTAAAA 3060  
3001 GCACATGGTTACATTTGGGAACTTTGTAGGATCATGGCCCTGTAGCCAGGTTTAAAA 3060  
3061 ACTGGACCTTTTAGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3120  
3061 ACTGGACCTTTTAGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3120  
3121 ATATGTTTTTCTTCCCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3180  
3121 ATATGTTTTTCTTCCCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3180  
3181 ATGTTATCTTTCTGATTCATGTAATCAATCTTTCTTGTGCAATTTTGGATGAGAT 3240  
3181 ATGTTATCTTTCTGATTCATGTAATCAATCTTTCTTGTGCAATTTTGGATGAGAT 3240  
3241 TTGAAGTTAAATGCAATTAAGAAAACCTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3300  
3241 TTGAAGTTAAATGCAATTAAGAAAACCTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3300  
3301 AAATATATACAGTAAATCCACATTTTCAATGATCTCTGGGAAATGATAGGATATGCAA 3360  
3301 AAATATATACAGTAAATCCACATTTTCAATGATCTCTGGGAAATGATAGGATATGCAA 3360  
3361 ATAAAGTTGAGTTTGTAGAGGGTAAACAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3420  
3361 ATAAAGTTGAGTTTGTAGAGGGTAAACAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 3420  
3421 GAAATTAACAAATTCAGAAATATATATCTGAAATTAATTTATAGGCTTTTCTCCACATTT 3480  
3421 GAAATTAACAAATTCAGAAATATATATCTGAAATTAATTTATAGGCTTTTCTCCACATTT 3480  
3481 AGACAAACCAACATATCTTAAAGCTTAAATTAATATATTTTCTTAAGGCTCAGTTGG 3540  
3481 AGACAAACCAACATATCTTAAAGCTTAAATTAATATATTTTCTTAAGGCTCAGTTGG 3540  
3541 GACATTAACCTTAAGAAAACATATCTTAAAGCCTTTTAAACCTTATTTTAGGACCTTT 3600  
3541 GACATTAACCTTAAGAAAACATATCTTAAAGCCTTTTAAACCTTATTTTAGGACCTTT 3600  
3601 TCGTGTGGGATGGGGCAAGGGTGGAGGTTTTAGAGAGTATATATCTCTTTAAAAA 3660  
3601 TCGTGTGGGATGGGGCAAGGGTGGAGGTTTTAGAGAGTATATATCTCTTTAAAAA 3660  
3661 AAAACAGAAAGAAAATATTTCTGAGCCTCATTAGCCCTATATCTCTTTAAAAA 3720  
3661 AAAACAGAAAGAAAATATTTCTGAGCCTCATTAGCCCTATATCTCTTTAAAAA 3720  
3721 TTTGTAGGGCCAGTTATCTGAGCAGATTTGCAATGTTTACCAAGATTTCTTAAAAATGAGT 3780  
3721 TTTGTAGGGCCAGTTATCTGAGCAGATTTGCAATGTTTACCAAGATTTCTTAAAAATGAGT 3780  
3781 GCAGATTCTGAATATAATACATTTTAAAAATTTTGGGAGTAGTATAATTTGTTGAGA 3840





QY 6001 TTAGCAATGGCTGCTGTTAGTTCTGGTAAATGTGCTAAAGTTGAATTTGTTCTGTTT 6060  
Db 6001 TTAGCAATGGCTGCTGTTAGTTCTGGTAAATGTGCTAAAGTTGAATTTGTTCTGTTT 6060  
QY 6061 TTCTCACACTGTGTGACGACCCATGTCTCAACACAGATAGTGTGTGATCACATAG 6120  
Db 6061 TTCTCACACTGTGTGACGACCCATGTCTCAACACAGATAGTGTGTGATCACATAG 6120  
QY 6121 ATCTACATAGTTGTGTCAGTTTGTGCTAAACCCCATAGGAGCTCTTTGGGATCAT 6180  
Db 6121 ATCTACATAGTTGTGTCAGTTTGTGCTAAACCCCATAGGAGCTCTTTGGGATCAT 6180  
QY 6181 GAAAGAGATCATGCAACAGCAATTTGGTGAAGGACACCTCAGATGCACTTTAGGCGCTT 6240  
Db 6181 GAAAGAGATCATGCAACAGCAATTTGGTGAAGGACACCTCAGATGCACTTTAGGCGCTT 6240  
QY 6241 TCTATGATGTTGTCAACCTCTGAGATGGAAGGCGAGTGTCTTTGATGTTATAGCCT 6300  
Db 6241 TCTATGATGTTGTCAACCTCTGAGATGGAAGGCGAGTGTCTTTGATGTTATAGCCT 6300  
QY 6301 AGAATGACACAGAACTATTGCTAATGATATAAACAACCTTCAATATATATTTGA 6360  
Db 6301 AGAATGACACAGAACTATTGCTAATGATATAAACAACCTTCAATATATATTTGA 6360  
QY 6361 ACAGATGAACAGAAATGTAATGTTTCTCTCAGAAACACTCTCTCAATATATATTTGA 6420  
Db 6361 ACAGATGAACAGAAATGTAATGTTTCTCTCAGAAACACTCTCTCAATATATATTTGA 6420  
QY 6421 TCATGCTGTAATGTAACCTTGGGCTACAACTCTTCAATGCTGTACAACTCTCTGCTC 6480  
Db 6421 TCATGCTGTAATGTAACCTTGGGCTACAACTCTTCAATGCTGTACAACTCTCTGCTC 6480  
QY 6481 ATTCAGTCTGATTTTATCCATAGAAAGAGCTACATTTAGTGTAAAGTGTACAT 6540  
Db 6481 ATTCAGTCTGATTTTATCCATAGAAAGAGCTACATTTAGTGTAAAGTGTACAT 6540  
QY 6541 ATATTTTATACGTGACTTAATTTGTCATTAACAACTTTTACACCAACCAATGTA 6600  
Db 6541 ATATTTTATACGTGACTTAATTTGTCATTAACAACTTTTACACCAACCAATGTA 6600  
QY 6601 CATGTGCACTGCAAAAGAGATCTCGGACATGCAAAATGTTTACCAAGAACCAACCGCT 6660  
Db 6601 CATGTGCACTGCAAAAGAGATCTCGGACATGCAAAATGTTTACCAAGAACCAACCGCT 6660  
QY 6661 TTGTCCCAAGGTGACTGTAATGTAAGTGAAGTGGCTTTTAAATAGGTGTGAGT 6720  
Db 6661 TTGTCCCAAGGTGACTGTAATGTAAGTGAAGTGGCTTTTAAATAGGTGTGAGT 6720  
QY 6721 GAAGAACATGCTGATGTTACTTAACAGCCCTTTGAATTTTAAACAACTGGGAATCCAT 6780  
Db 6721 GAAGAACATGCTGATGTTACTTAACAGCCCTTTGAATTTTAAACAACTGGGAATCCAT 6780  
QY 6781 AGGAACCGATGCTATCATCTGAACATAGCTGGAGCTGGAATGTAATTTTACG 6840  
Db 6781 AGGAACCGATGCTATCATCTGAACATAGCTGGAGCTGGAATGTAATTTTACG 6840  
QY 6841 AATGAAAAGTGTGTGACTAGTACTTAAAGTGTCTAATGATAAGTTTGAAGTCAA 6900  
Db 6841 AATGAAAAGTGTGTGACTAGTACTTAAAGTGTCTAATGATAAGTTTGAAGTCAA 6900  
QY 6901 AATGAAAAGAAAAAATCTGCAATCCAGGCGAATTTGATATTTTATGCAATTTAA 6960  
Db 6901 AATGAAAAGAAAAAATCTGCAATCCAGGCGAATTTGATATTTTATGCAATTTAA 6960  
QY 6961 AATGCTATTTCTGTAATTTGGGAAATCAAGTGGCTTATCATGATATCGTGTACTTAA 7020  
Db 6961 AATGCTATTTCTGTAATTTGGGAAATCAAGTGGCTTATCATGATATCGTGTACTTAA 7020  
QY 7021 ATGTAATTCACAACTACTGTTGTAATTTGTAATATATAGACAAAGATCATATTTTGT 7080  
Db 7021 ATGTAATTCACAACTACTGTTGTAATTTGTAATATATAGACAAAGATCATATTTTGT 7080

QY 7081 GTGTATTAAGCTCTGTAAATAGCAATCAATTATGAAGCTGCACTGATACTACATTTT 7140  
Db 7081 GTGTATTAAGCTCTGTAAATAGCAATCAATTATGAAGCTGCACTGATACTACATTTT 7140  
QY 7141 AAACATTCACATCCAAAGAGCAGACTATTATTATGTCATATACCAATTTAAATATTA 7200  
Db 7141 AAACATTCACATCCAAAGAGCAGACTATTATTATGTCATATACCAATTTAAATATTA 7200  
QY 7201 ATTTGCTGCTAAATTAATTAATAGTACTGCACTTTCTTGTCCTACAGTGTATGTTTC 7260  
Db 7201 ATTTGCTGCTAAATTAATTAATAGTACTGCACTTTCTTGTCCTACAGTGTATGTTTC 7260  
QY 7261 TGTAAAGTAAGTATGTAATTTCCACAAATATATGAATAAATCTCGTGCC 7313  
Db 7261 TGTAAAGTAAGTATGTAATTTCCACAAATATATGAATAAATCTCGTGCC 7313

RESULT 2  
AAV18481  
ID AAV18481 standard; cDNA; 2334 BP.  
XX  
AC AAV18481;  
XX  
DT 18-AUG-1998 (first entry)  
XX  
DE Human zinc finger protein (hZAC).  
XX  
KW Corticotroph; pituitary; BOP1; tumour suppressor; constitutive;  
KW inducible; Alzheimer's disease; nuclear transcription factor; apoptosis;  
KW cell cycle; neuronal disorder; human zinc finger; hZAC; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 803..2194  
FT /\*tag= a  
FT /product= "hZAC"  
XX  
PN WO9813489-A1.  
XX  
PD 02-APR-1998.  
XX  
PF 22-SEP-1997; 97WO-EP05198.  
XX  
PR 23-SEP-1996; 96US-0718661.  
XX  
PA (CNRS ) CENT NAT RECH SCI.  
PA (PLAC ) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.  
XX  
PI Journot L, Spengler D;  
XX  
DR WPI; 1998-230701/20.  
XX  
DR P-PSDB; AAW48761.  
XX  
PT New isolated tumour suppressor gene - useful for developing products  
PT for use in diagnosis and treatment of tumour(s) or neuronal  
PT disorder(s)  
XX  
PS Claim 17; Pages 82-85; 118pp; English.  
XX  
CC The present sequence represents the human zinc finger protein  
CC (hZAC) cDNA isolated from the human pituitary gland cDNA library  
CC using BOP1 cDNA (AAV18480) as a probe. Therefore the hZAC protein  
CC is the human homologue of BOP1 (AAW48761). The hZAC protein displays  
CC a tumour suppressing activity when it was constitutively and inducibly  
CC expressed in tumour cells. The hZAC cDNA and the protein it encodes  
CC are claimed to be useful in the preparation of therapeutic compositions,  
CC useful for treating, preventing or delaying the recurrence of a tumour  
CC or neuronal disorders, e.g. genetic diseases or acquired degenerative  
CC encephalopathies such as Alzheimer's disease. The hZAC protein is  
CC also claimed to be able to induce apoptosis resulting in inhibition  
CC of tumour cell growth, to suppress tumour formation, to induce G1  
CC arrest of the cell cycle and to act as nuclear transcription factor.

XX	PD	16-MAY-2002.	
SQ	XX	08-NOV-2001; 2001WO-US46816.	
	XX	09-NOV-2000; 2000US-247457P.	
	XX	(BGM ) BRIGHAM & WOMENS HOSPITAL INC.	
	XX	Lee RT;	
	XX	WPI; 2002-590446/63.	
	DR	Diagnosing cardiovascular conditions e.g. stroke, myocardial infarction	
	XX	in a subject, by detecting expression of a nucleic acid molecule such	
	PT	as Fit-1, CD44, Lot-1, AA892598 and Mrg-1 in biological sample from	
	PT	subject -	
	XX	Disclosure; Page 107-108; 113pp; English.	
	XX	The invention relates to diagnosing a cardiovascular condition	
	CC	characterised by aberrant expression of a nucleic acid molecule (I) such	
	CC	as Fit-1, CD44, Lot-1, AA892598 and Mrg-1 or its expression product. One	
	CC	method involves contacting a biological sample from a subject with an	
	CC	agent which specifically binds to (I), its expression product or a	
	CC	fragment of an expression product and measuring the amount of bound	
	CC	agent. The method is useful for diagnosing a cardiovascular condition	
	CC	such as myocardial infarction, stroke, arteriosclerosis, heart failure,	
	CC	and cardiac hypertrophy. Methods useful for determining the stage of	
	CC	cardiovascular condition and pharmaceutical compositions for treating the	
	CC	above cardiovascular conditions are also provided. The present sequence	
	CC	represents a human ZAC zinger finger protein cDNA.	
	XX	Sequence 2828 BP; 701 A; 722 C; 651 G; 754 T; 0 other;	
	SQ	Query Match 4.8%; Score 351.6; DB 24; Length 2828;	
		Best Local Similarity 69.8%; Pred. No. 6.3e-57;	
		Matches 490; Conservative 0; Mismatches 209; Indels 3; Gaps 1;	
QY	543	GAACGTTAAGCGTGTGAAACCAACCAAGCAAGCCATTTCTTGGCAACTGTGTGACAA 602	
Db	1269	GAAGAGAGGTGTGAGAGCAAGCCATGCGCCATGTCCTGCGCAGTTATGTGGCAA 1328	
QY	603	GGCCTTTAAGTGTGAGAAATTAAGTTTCACTCTCTACTCTCACACGAGAGAGGCC 662	
Db	1329	GACGTTCTCCTCAGCTGTGAGAAATTCAGATTCAATTTATCCACCTCCAGGGAGCGCC 1388	
QY	663	CTACAGTGCATACCAAGACTGCACCAAGGCTTTTCTTAAGTACAAATTCACAAAG 722	
Db	1389	GTACAGTGTGTGAGGCTGACTGTGCAAGGCTTTGTTTCCAGATATAAATTGATGAG 1448	
QY	723	GCATATGGCTACTACTCTCTGAGAAACCCCAAGTAAATTTTGTGAGAAATGTT 782	
Db	1449	GCATATGGCTACTACTCTCTGAGAAATTTCTCCAGAGTCTCCTGCTGAGAGAGCTT 1508	
QY	783	TCACCGGAAGATCATCTGAGATACCTTCCATACACAGCACCCTTACCAAGAGAGCTT 842	
Db	1509	CAACCGGAAGAGCACCCTGAAAAACCACTCCAGACCCACACCCCAAAATATGGCTT 1568	
QY	843	TAAGTCGGAAGATGTGCAAGTACTCAATACCAAGCTTGATTTTAAACCTCACTTGGC 902	
Db	1569	TGGGTGTGAGGAGTGTGGAGAAAGTACACACCATGCTGGGCTGTGCTTAAAGACCTTGC 1628	
QY	903	CTTGCATGCCCAACCAAGTGTGTGACCTCACCCTGTAAGTATGTTTGAACCTTTTGAAG 962	
Db	1629	CCTCATGCGCCAGCAGTGGGACCTCACCCTGTTGGGCTGTGCTGCTTAAAGACCTTGC 1688	
QY	963	CACGGAGTGTCTTGTGAGCAGCTTAATCTCATGC---AGGCAAGTCTGTGGTGGGT 1019	
Db	1689	CACCGAGTGTCTTGTGAGCAGCTTAATCTCATGC---AGGCAAGTCTGTGGTGGGT 1748	
QY	1020	TAAAGAAAAAGCAGCAGTGTGCAACATTTGTGATCGCCGGTCTTACACCCGAAAGATGT 1079	
Db	1749	CAAGGAAAAAGCAGCAGTGTGCAACATTTGTGATCGCCGGTCTTACACCCGAAAGATGT 1808	

XX	PD	Sequence 2334 BP; 569 A; 636 C; 550 G; 579 T; 0 other;	
SQ	XX	Query Match 4.8%; Score 351.6; DB 19; Length 2334;	
	XX	Best Local Similarity 69.8%; Pred. No. 6e-57;	
	XX	Matches 490; Conservative 0; Mismatches 209; Indels 3; Gaps 1;	
QY	543	GAACGTTAAGCGTGTGAAACCAACCAAGCAAGCCATTTCTTGGCAACTGTGTGACAA 602	
Db	775	GAAGAGAGGTGTGAGAGCAAGCCATGCGCCATGTCCTGCGCAGTTATGTGGCAA 834	
QY	603	GGCCTTTAAGTGTGAGAAATTAAGTTTCACTCTCTCTCACACAGAGAGAGGCC 662	
Db	835	GACGTTCTCCTCAGCTGTGAGAAATTCAGATTCAATTTCCACTCCAGGAGCGGC 894	
QY	663	CTACAGTGCATACCAAGACTGCACCAAGGCTTTGTTTCTAAGTACAAATTCACAAAG 722	
Db	895	GTACAGTGTGTGAGGCTGACTGTGGCAAGCTTTGTTTCCAGATATAAATTTGATGAG 954	
QY	723	GCATATGGCTACTACTCTCTGAGAAACCCCAAGTAAATTTTGTGAGAAATGTT 782	
Db	955	GCATATGGCTACTACTCTCTGAGAAATTTCACTCAGTGTGCTCTGTGAGAGAGCTT 1014	
QY	783	TCACCGGAAGATCATCTGAGAAATCAGCTTCCATACACAGCACCCTTAAACAGAGAGCTT 842	
Db	1015	CAACCGGAAGAGCACCCTGAAACCCACCTTCCAGCCAGCACCACCAAAATGGCCTT 1074	
QY	843	TAAAGTGCAGAAATGTGGCAAGAACTACATACCAAGCTTGGATTTTAAAGTCACTTGGC 902	
Db	1075	TGGGTGTGAGGAGTGTGGGAAGATACACACCATGCTGGGCTATAAGAGGACCTGGC 1134	
QY	903	CTTGATGCGCGAACAAGTGTGAGTGCCTCCTGTAGGATGTTTGCAAACTTTTGAAG 962	
Db	1135	CCTCCATGCGCGCAGAGTGGGACCTTCACTGTGGGCTGTGCGCTGGAGCTAGGAG 1194	
QY	963	CACGGAGTCTCTTGTGAGCAGCTTAAATCTCATGC---AGGCAAGTGTGCTGTGGGT 1019	
Db	1195	CACGGAGTCTCTTGTGAGCAGCTTAAAGCCCTTCCAGCCATCGGAAGAGAGCCCTTAGCGAAC 1254	
QY	1020	TAAAGAAAAAGCAGCAGTGTGCAACATTTGTGATCGCGGCTTTCACACCGAAGAGTGT 1079	
Db	1255	CAAGGAAAGAGCACCAGTGTGAGGCTTGTGAAAGATGCTTCTACACCGGAGAGATGT 1314	
QY	1080	CGGAGACACATGTGTGTGACATGGAAGAGAGCTTCTCTGTGAGTATGTGACAA 1139	
Db	1315	GCACGCGCCTGTGTGTCCACAGAGATGCAAGGCTTCTGTGCGAGTCTGTGCGCCA 1374	
QY	1140	GAGATTTGGCGAAGAGATCAGCTGACTCCGATATGAAGAGAGTCACAATCAAGAGCT 1199	
Db	1375	GAGATTTGGCGCAAGAGATCAGCTTCCCGCATACCAAGAGAGCACCCTCAGAGAGCT 1434	
QY	1200	TCTGAAGTCAAAACAGAACAGTGTGATTTCTTGTGACCCATT 1241	
Db	1435	GATGAAGAGAGCTTGCAGACCGGAGACCTTCTGAGCACCTT 1476	

RESULT 3

ABN86525

ID ABN86525 standard; cDNA; 2828 BP.

XX AC ABN86525;

XX 21-OCT-2002 (first entry)

DT Nucleotide sequence of human ZAC zinger finger protein cDNA.

DE Cardiovascular; Fit-1; CD44; Lot-1; AA892598; Mrg-1; cardiant; human;

KW cerebroprotective; antiarteriosclerotic; ZAC; zinger finger protein; ss.

XX OS Homo sapiens.

XX PN WO200238794-A2.

XX



## RESULT 6

ABN86508  
ID: ABN86508 standard: cDNA: 5028 BP.

AA  
AC ABN86508;

XX 21-OCT-2002 (first entry)

XX : rat tot-1 polypeptide encoding cDNA.

XX XX  
Cardiovascular; Fit-1; CD44; Lot-1; AA892598; Mrg-1; cardiant; rat;  
KW KW  
obstructive; arteriosclerotic; gene therapy; gene; ss.

XX  
Rattus norvegicus. OS

XX	Key	Location/Qualifiers
FH		
FT	CDS	548..2299

```

FF      /*tag= a
FF      /product= "Lot-1"

```

XX PN WO200238794-A2.

XX

## RESULT 5

ABN86524

XX  
AC ABN86524;

XX  
DT 21-OCT-2002 (first entry)

XX Nucleotide sequence of human Lot-1 cDNA.

XX Cardiovascular; Pit-1; CD44; Lot-1; AA892598; Mrg-1; cardiand; human;  
KW cerebromprotective; antiarteriosclerotic; ss.  
KW

XX Homo sapiens.

XX  
PN WO200238794-A2.XX  
16-MAY-2002[illegible]

XX  
PR 09-NOV-2000; 2000US-247457P.

XX  
XX  
(BGHM ) BRIGHAM & WOMENS HOSPITAL INC. PA

XX	DT	TC	PT.
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100

[illegible]

XX Diagnosing cardiovascular conditions e.g. stroke, myocardial infarction  
PT in a subject, by detecting expression of a nucleic acid molecule such  
PT as Ftt-1, CD44, Lot-1, AA892598 and Wrg-1 in biological sample from  
PT subject -  
PT

XX PS Disclosure: Page 105-107; 113pp; English.

xx The invention relates to diagnosing a cardiovascular condition  
cc characterised by aberrant expression of a nucleic acid molecule (I) such  
cc as Fit-1, CD44, Lot-1, AA82598 and Mrg-1 or its expression product. One  
cc method involves contacting a biological sample from a subject with an  
cc agent which specifically binds to (I), its expression product or a  
cc fragment of an expression product and measuring the amount of bound  
cc agent. The method is useful for diagnosing a cardiovascular condition  
cc such as myocardial infarction, stroke, arteriosclerosis, heart failure,  
cc and cardiac hypertrophy. Methods useful for determining the stage of  
cc cardiovascular condition and pharmaceutical compositions for treating the  
cc above cardiovascular conditions are also provided. The present sequence  
cc represents a human Lot-1 cDNA.

XX 1243 T: 1 other:

Query Match	4.8%	Score 350;	DB 24;	Length 4632;
Best Local Similarity	69.7%	Pred. No. 1.4e-56;		



Query Match	4.0%; Score 294.4; DB 19; Length 2790;
Best Local Similarity	64.6%; Pred. No. 4.2e-46;
Matches	455; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
557	GTGAAACCAACCAAGAAAAAATCTTCCTGCCAACTGTGTGACAAGGCCCTTTAAACAGTG 616
528	GAGAACCAAGAGGCATGGCTCCATTCGCTGTCAAATATGTGGCAAGTCTCTTCGTACCC 587
617	TTGGAATTAATAAGGTTCACTCCTCACTCTCAACAGGAGAGAGGCCCTTACAAGTGCATAC 676
588	TGGAGAAGTTTCAACATTACAAATATTCCCACTCCAGGAGCGGCCATTCAAGTGTCTCGA 647
677	AACAAGACTGCACGAAGGCCCTTTGTTCTTAAGTACAAATTACAAGGACCATGGTACTCTC 736
648	AGGCTGAGTGTGGCAAGCCTTCGTTCTCCAAGTATAAGCTGATGAGACACATGGCCACAC 707
737	ATTCTCTCGAAGAACCCACAGTGTAAATTATTGTGAGAAATGTTTCAACCGGAAAGATC 796
708	ACTGCCACAGAAGATTCAACAGTGTACTACTGTGAGAGACATTCACCCGAGGAGACC 767
797	ATCTGAAGAATCACCTCCATACACAGCACCCCTTAACAAAGAGACGTTTAAAGTGCGAAGAT 856
768	ACCTGAGAGACCACTCCAGAGCCACGATCCCAACAAGATCTCTACGCGTGTGACGATT 827
857	GTGGCAAGAACTACAATACCAAGCTTGATTTAAAGTCACATTGGCCTTGCAATGCCGAA 916
828	GCGGCAAGAAGTACCACACCAATGCTGGGCTACAAGAGGCACCTGGCCCTTGCACTGGCGA 887
917	CAAGTGGTGACCTCACTTAAGGTATGTTTGCAGAACTTTTCAAGACACGGAGTGCTCTC 976
888	GGAATGGCGATCTCACTGTGGGTGTGCACCTGGAGCTGGGAGCACCGAGTCTCTGC 947
977	TGGAGCACCTTAAATCTATGC---AGGCAAGTCGTTCTGGTGGGTTTAAAGAAAAAAGC 1033
948	TGGAACACCTCAAGTCTACGCGGAAGAAAGGCCAACAGGCACCCAGGAGAGAAAT 1007
1034	ACCAGTGCAGCAATTGTGATGCGCGTCTACACCGGAAGAGATGTCGGAGACACATGG 1093
1008	ACCAGTGCACCACTGTGATAGATGTTTACACCCGGAAAGATGTGCGTTCGCCACCTCG 1067
1094	TGTTGCACACTGGAGAAAGGACTTCCTCTGTCAAGTATGTGCACAGAGATTTGGCGCAA 1153
1068	TGTTCCACACAGGATCAAGGACTTCCTGTGTGAGTCTGTGCCCCAAGATTTGGCGCA 1127
1154	AGGATCACTGTACTGCACATATGAAGAAGAGTCACAATCAAGAGCTTCTGAAGGTCAAAA 1213
1128	AGAACCCTCACTCGTCACACCAAGAGACCCACTCCAGGAGCTGATGCAAGAGAATA 1187
1214	CAGAACCGTGGATTTCCTTGACCCATTTACTGGCAATGHTCT 1257
1188	TGCAGCAGGAGATTAACAGAGCAATTTCCAACTCATTTGCGCCT 1231

RESULT 8	
ABN86526	
ID	ABN86526 standard; cDNA; 3975 BP.
XX	
XX	
AC	ABN86526;
XX	
XX	21-OCT-2002 (first entry)
DT	
XX	
XX	
DE	Nucleotide sequence of mouse ZAC zinger finger protein cDNA.
XX	
XX	Cardiovascular; Fit-1; CD44; Lot-1; AAG92598; Mrp-1; cardiant; mouse;
KW	cerebroprotective; antiarteriosclerotic; ZAC; zinger finger protein; ss.
KW	
XX	
OS	Mus musculus.
OS	
XX	
XX	WO200238794-A2.
PN	
XX	
XX	
PD	16-MAY-2002.
XX	
XX	
PP	08-NOV-2001; 2001WO-US46815.

09-NOV-2000; 2000US-247457P.  
(BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
Lee RT;  
WPI; 2002-590446/63.  
Diagnosing cardiovascular conditions e.g. stroke, myocardial infarction in a subject, by detecting expression of a nucleic acid molecule such as Fit-1, CD44, Lot-1, AA892588 and Mrg-1 in biological sample from subject -  
disclosure: Page 108-109; 113pp: English.

The invention relates to diagnosing a cardiovascular condition characterized by aberrant expression of a nucleic acid molecule (I) such as Fit-1, CD44, Lot-1, AA892598 and Mrg-1 or its expression product. One method involves contacting a biological sample from a subject with an agent which specifically binds to (I), its expression product or a fragment of an expression product and measuring the amount of bound agent. The method is useful for diagnosing a cardiovascular condition such as myocardial infarction, stroke, arteriosclerosis, heart failure, and cardiac hypertrophy. Methods useful for determining the stage of cardiovascular condition and pharmaceutical compositions for treating the above cardiovascular conditions are also provided. The present sequence represents a mouse ZAC zinger finger protein cDNA.

XX	Sequence	3975 BP;	1010 A;	1001 C;	880 G;	1084 T;	0 other;
Query	Match	4.0%;	Score 294.4;	DB 24;	Length 3975;		
Best	Local Similarity	64.6%;	Pred. No. 4.5e-46;				
Matches	455; Conservative	0;	Mismatches 246;	Indels 3;	Gaps 1;		
QY	557	GTGAACCAACCAAGAAAACCTTTCCTGTGCCAACTGTGTGTGCAAGAGCGCTTTAAACAGTG	616				
DB	235	GAGAACCAAGGCGATGGCTCCATTCGGTGTCAAAATGTGCAAGTCTTTCGTACACC	294				
QY	617	TTGAGAAATTAAGGTTCACTCTACTCTCACACGAGAGAGCGCCTCAAGTGCATAC	676				
DB	295	TGAGAAAGTTCACCAATTCACAAATATTCACCTCCAGGAGCGCCCATTCAGTGTTCGA	354				
QY	677	AACAAGACTGCACAGAGCGCTTTGTTCTTAAGTACAAATTCAAAGGCACATGCTACTC	736				
DB	355	AGGCTGAGTGTGGCAAGCGCTTCGTCTCCAAATTAAGCTGATGAGACATGGCCACAC	414				
QY	737	ATTCTCTGTGAGAAAACCCACAAGTGTAAATTAATGTGAGAAAATGTTTTCACCGGAAAGATC	796				
DB	415	ACTCGCCACAGAGATTCACCAAGTGTACTCACTGTGAGAAAGACATTCACCGGAAAGGACC	474				
QY	797	ATCTGAAGATACCTCCATFACACAGCCGTTAAACAAGAGACGTTTAAAGTGGCAAGAT	856				
DB	475	ACCTGAAGAACCACCTCCAGACCCACCATCCCAACAGATCTCTACGCGTGTGACGATT	534				
QY	857	GTGCGAAGAACTACAATACCAAGCTTCGATTTAAACGTTCATTTGGCCCTTGATCCCGCAA	916				
DB	535	GCGCGAAGAGTACACACCATGCTGGGCTACAGAGGCACCTGGCCCTGCACCTCGCGCA	594				
QY	917	CAAGTGGTGACCTCACCTGTAAAGTATGTTTGTGAAACTTTTGAAGACAGGAGTGTCTC	976				
DB	595	GCAATGGCATCTCACCTGTGGGCTGTGCACTTGGAGCTGGGAGACACGAGTCTCTGC	654				
QY	977	TGGAGCACCTTAATCTCATGSC---AGCGAAGTCTGTCTGTGGGGTTTAAAGAAAAAAGC	1033				
DB	655	TGGACCACTCAAGTCTCACGCGAAGAAAGGCCAACCGACCCAGGAGAGAAGAAAT	714				
QY	1034	ACCAAGTGCAGCAATGTGATCGCGGTTCTACACCGGAAGAGTGTCCGAGACACATGG	1093				
DB	715	ACCAAGTGCAGCACTGTGATAGATGCTTCTACACCGGAAAGATGTGGCTGCCACCTGG	774				
QY	1094	TGTTGCACACTGGAGAAAGACTTTCCTCTGTCTAGTATTTGTGCAAGAGATTTGGGCGAA	1153				



Db 775 TGGTCCACAGGATGCAAGGACTTCCGTGTCTAGTCTGTGCCACAGAGATTGGGGCGCA 834  
 QY 1154 AGGATCCTGACTCGACATATGAAGAAGAGTCAATCAAGAGCTTCTGAAGTCAAAA 1213  
 Db 835 AAGACCACTCTCACTCGTTCACACCAAGAGAGCCCACTCCAGAGAGCTGATGCAAGAGAATA 894  
 QY 1214 CAGAACCAAGTGGATTCTCTTGACCCATTTACTCTGCAATGTCT 1257  
 Db 895 TGCAGGAGGAGATTACAGAGCAATTTCCAACTCAITGGCGCT 938

# RESULT 9 AAV29269

ID AAV29269 standard; cDNA; 2561 BP.

AC AAV29269;

XX 21-AUG-1998 (first entry)

DE Nucleotide sequence of human PLAG2.

XX Human PLAG2 gene; PLAG2; tumorigenesis gene; T-gene; PLAG1; CTNNB1;  
 KW antibody; benign tumour; malignant tumour; leukaemia; lymphoma;  
 KW cancer; inhibition; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
 FH 177..1416  
 FT CDS /tag= a  
 FT /product= "PLAG2 protein"

XX EP825198-A1.

XX 25-FEB-1998.

XX 17-JAN-1997; 97BP-0200130.

XX 22-AUG-1996; 96BP-0202339.

XX (KULE-) KU LEUVEN RES & DEV.

XX (UYGO-) UNIV GOETTERBORGS HOLDINGBOLAGET AB.

XX (LEUV-) LEUVEN RES & DEV.

XX Kas KP, Stenman KGD, Van De Ven WJM, Voz ML;

XX WPI; 1998-132252/13.

XX P-PSDB; AAW37949.

XX New tumorigenesis T-genes and proteins - useful for, e.g. preparing  
 PT antibodies for clinically diagnosing cells having non-physiological  
 PT proliferative capacity such as lipoblastomas

XX Claim 4; Fig 8; 71pp; English.

XX This is the nucleotide sequence of the human PLAG2 gene. It is a  
 CC tumorigenesis gene (T-gene), which is isolated in the form of PLAG1,  
 CC PLAG2, and CTNNB1 genes. Their proteins can be used as a starting  
 CC point for preparing antibodies for clinically/medically diagnosing  
 CC cells having a non-physiological proliferative capacity as compared  
 CC to wild type cells, where the former cells are selected from both  
 CC benign and malignant tumours, as well as leukaemia and lymphomas.  
 CC Derivatives of the T-gene are also used in the diagnosis and  
 CC preparation of therapeutic compositions for the treatment of cancers,  
 CC such as nucleic acid derivatives, and antibodies. The T-gene may be  
 CC used as a starting point for designing suitable expression-modulating  
 CC compounds or techniques for the treatment of non-physiological  
 CC proliferation phenomena in humans or animals. Expression inhibitors  
 CC of the T-gene can be used in the treatment of diseases involving  
 CC benign or malignant tumours.

XX Sequence 2561 BP; 704 A; 624 C; 516 G; 716 T; 1 other;

Query Match 3.8%; Score 277.6; DB 19; Length 2561;  
 Best Local Similarity 71.9%; Pred. No. 6.2e-43;  
 Matches 377; Conservative 0; Mismatches 144; Indels 3; Gaps 1;  
 QY 721 AGGCACATGCTACTCTCTCTGAGAAAACCCACCAAGTGAATATTGTGAGAAAATG 780  
 Db 171 AGGCATATGCTACCCCAATCTCCCCAGAAATCTCACCAGTGTGCTCACTGTGAGAGACG 230  
 QY 781 TTTTACCGGAAAGATCATCTGAAGAATCACCTCATACACAGCCCTTACAAAGAGACG 840  
 Db 231 TTCAACCGGAAAGACCACTGAAACACCACTCCAGACCACCCCAACCAAAATGGCC 290  
 QY 841 TTTAAGTCCGAAGAATGTGCAAGAACTACAATACCAAGCTTGGATTATAACCGTCACTTG 900  
 Db 291 TTTGGGTGTGAGGAGTGTGGGAAGAAGTACAACACCATGCTGGGCTATAAGAGCCACTG 350  
 QY 901 GCCTTGATGCGCGCAACAAGTGTGACCTCACTGTAAGTATGTTTGCAAACTTTTGA 960  
 Db 351 GCCTTCATGCGCGCCAGCAGTGGGACCTCACCTGTGGGTCTGTGCCCTGGAGCTAGGG 410  
 QY 961 AGCAGCGGAGTGTCTCTGAGGACACCTTAAATCTCATGC--AGGCAAGTCTGTGTGGG 1017  
 Db 411 AGCAGCGAGTGTCTACTGSAACCACTCAAGCCCATGCGGAAGAGAGGCCCTTAGCGGA 470  
 QY 1018 GTTAAAGAAAAAAGCACCAAGTCCGCAACATTTGTATCGCGGTTCTACACCCGAAAGGAT 1077  
 Db 471 ACCAAGGAAAAAGACCACTGCGGACCACTGTGAAGATGCTTCTACACCCGAAAGGAT 530  
 QY 1078 GTCCGAGACACATGTTGTGTGCACACTGCAAGAAAGGACTTCTCTGTGAGTATTGTGCA 1137  
 Db 531 GTGCGAGCCCACTGTGTGTCCACACAGGATGCAAGGACTTCTGTGCGAGTCTCTGTGCC 590  
 QY 1138 CAGAGATTTGGCGGAAAGGATCACCTGACTCCACATATGAAGAAGTCAACAATCAAGAG 1197  
 Db 591 CAGAGATTTGGCGGCAAGGTTCACTCACCAGCATACCAAGAGACCCATCAGAGAG 650  
 QY 1198 CTTCTGAAGGTTCAAAACAGAACCAAGTGGATTTCCTTGGACCCATT 1241  
 Db 651 CTGATGAAGAGAGGCTTGACAGCGGAGACCTTCTGAGCACCTT 694

## RESULT 10

ABT11028  
 ID ABT11028 standard; cDNA; 2561 BP.

XX AC ABT11028;

XX XX

DT 04-DEC-2002 (first entry)

XX Human breast cancer associated coding sequence SEQ ID NO: 1162.

XX Human; breast specific gene; breast cancer; differential expression;  
 KW cytostatic; gene therapy; gene; ss.

XX Homo sapiens.

XX WO200259271-A2.

XX 01-AUG-2002.

XX 25-JAN-2002; 2002WO-US02176.

XX 25-JAN-2001; 2001US-263757P.

XX 25-APR-2001; 2001US-286090P.

XX 23-MAY-2001; 2001US-292517P.

XX (GENE-) GENE LOGIC INC.

XX Orr MS, Nation M, Diggans JC, Zeng W;

XX WPI; 2002-674803/72.

XX Diagnosing breast cancer in a patient comprises detecting the level of

PT gene expression in cell or tissue samples, where a differential gene  
PT expression is indicative of breast cancer -  
XX  
XX  
PS Claim 1; SEQ ID NO 1162; 260pp + Sequence Listing; English.  
XX  
XX The present invention relates to methods of diagnosing breast cancer in a  
CC patient, which comprise detecting the level of expression in a tissue  
CC sample of two or more genes selected from those shown in AB09867-  
CC AB01112, where a differential expression of the genes indicates breast  
CC cancer. The methods are useful in diagnosing, treating, detecting the  
CC progression, and in monitoring treatment of breast cancer in patients.  
CC The methods are also useful as a screening tool for agents that modulate  
CC the onset or progression of breast cancer. The breast cancer genes may be  
CC used as diagnostic markers for the prediction or identification of the  
CC malignant state of breast tissue, for confirming the type and progression  
CC of cancer, and for drug screening and assays. The present sequence is a  
CC coding sequence of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2561 BP; 704 A; 624 C; 516 G; 717 T; 0 other;

Query Match 3.8%; Score 277.6; DB 24; Length 2561;  
Best Local Similarity 71.9%; Pred. No. 6.2e-43; Indels 3; Gaps 1;  
Matches 377; Conservative 0; Mismatches 144;  
QY 721 AGGCATGCTACTCATCTCTCTGAGAAACCCACAGTGTAAATTATTGTGAGAAATG 780  
Db |||||  
QY 171 AGGCATGCTACTCATCTCTCTGAGAAACCCACAGTGTAAATTATTGTGAGAAATG 230  
Db |||||  
QY 781 TTTCACCGGAAAGACCATCTCTGAGAAACCCACAGTGTAAATTATTGTGAGAAATG 840  
Db |||||  
QY 231 TTTCACCGGAAAGACCATCTCTGAGAAACCCACAGTGTAAATTATTGTGAGAAATG 290  
Db |||||  
QY 841 TTTCACCGGAAAGACCATCTCTGAGAAACCCACAGTGTAAATTATTGTGAGAAATG 900  
Db |||||  
QY 291 TTTCACCGGAAAGACCATCTCTGAGAAACCCACAGTGTAAATTATTGTGAGAAATG 350  
Db |||||  
QY 901 GCCTTGATGCGGCAACAGTGTGACCTCACCTGTAGGATGATTTGTGCAAACTTTTGA 960  
Db |||||  
QY 351 GCCTTGATGCGGCAACAGTGTGACCTCACCTGTAGGATGATTTGTGCAAACTTTTGA 410  
Db |||||  
QY 961 AGCACGGAGTGTCTTGTGAGACCTTAAATCTCATGC---AGCAAGTCTCTGTGGG 1017  
Db |||||  
QY 411 AGCACGGAGTGTCTTGTGAGACCTTAAATCTCATGC---AGCAAGTCTCTGTGGG 470  
Db |||||  
QY 1018 GTTAAAGAAAGAAAGACCATCTCTGAGAAACCCACAGTGTAAATTATTGTGAGAAATG 1077  
Db |||||  
QY 471 ACCAAGGAAAGAAAGACCATCTCTGAGAAACCCACAGTGTAAATTATTGTGAGAAATG 530  
Db |||||  
QY 1078 GTCCGGAGACATGCTGCTGACACCTGAGAAAGAGACTTCTCTGTGATTTGTGCA 1137  
Db |||||  
QY 531 GTGAGCCGACCTGCTGCTGACACAGATGCAAGGACTTCTCTGTGAGTCTGTGCC 590  
Db |||||  
QY 1138 CAGAGATTGGCGGAAAGATCCTGATCTGATATGAGAGAGTCAATCAAG 1197  
Db |||||  
QY 591 CAGAGATTGGCGGAAAGATCCTGATCTGATATGAGAGAGTCAATCAAG 650  
Db |||||  
QY 1198 CTTCTGAGAGTCAAAACAGAACCATCTGAGTATTTCTTTGACCCATT 1241  
Db |||||  
QY 651 CTGATGAAAGAGAGCTTGCAGACCGGAGACCTTCTGAGGACCTT 694  
Db |||||

RESULT 11  
AAZ33549  
ID AAZ33549 standard; cDNA; 2738 BP.  
XX  
XX AC AAZ33549;  
XX  
XX 08-DEC-1999 (first entry)  
XX  
XX Human breast tumour-associated EST 9.

XX  
KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;  
KW medicaments; gene therapy; treatment; fat metabolism; ss.  
XX  
XX Homo sapiens.  
XX  
XX DE19813835-A1.  
XX  
XX 23-SEP-1999.  
XX  
XX 20-MAR-1998; 98DE-1013835.  
XX  
XX 20-MAR-1998; 98DE-1013835.  
XX  
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
XX  
XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;  
XX  
XX WPI; 1999-528979/45.  
XX  
XX P-PSDB; AAY48468.  
XX  
XX Human nucleic acid sequences and protein products from normal breast  
XX tissue, useful for breast cancer therapy -  
XX  
XX Claim 3; 96-97; 206pp; German.  
XX  
XX This invention describes novel human nucleic acid sequences from normal  
XX breast tissue which have cytostatic activity. The nucleic acid sequences  
XX can be used to produce and isolate full-length gene sequences. They can  
XX be used to express proteins, which can be used as tools to find an  
XX activity against breast cancer. The sequences can be used in sense or  
XX antisense form. They are especially useful for medicaments for gene  
XX therapy to treat breast cancer and for treating illnesses associated  
XX with fat metabolism. AAZ33541-233610 represent expressed sequence tags  
XX described in the method of the invention.  
XX  
XX Sequence 2738 BP; 771 A; 661 C; 546 G; 760 T; 0 other;

Query Match 3.8%; Score 276; DB 20; Length 2738;  
Best Local Similarity 71.2%; Pred. No. 1.3e-42;  
Matches 379; Conservative 0; Mismatches 150; Indels 3; Gaps 1;  
QY 713 AATTACAAAGGCATGCTACTCATCTCTCTGAGAAACCCACAGTGTAAATTATTGTG 772  
Db |||||  
QY 324 AATCAGTAGGCAAAATTGGCTACCCATTCTCCCGAAGATCTCACCAGTGTCTCAGTGTG 383  
Db |||||  
QY 773 AGAAATGTTTTCACCGGAAAGATCATCTGAAGATCATCTCCATPACACAGCCTTAACA 832  
Db |||||  
QY 384 AGAAGAGCTTCAACCGGAAAGACCATCTGAAACACCTCCAGACCCACGACCCCAACA 443  
Db |||||  
QY 833 AAGAGAGCTTTAAGTGGCAAGAAATGTGGCAAGAACTCAATPACCAAGCTTGGATTAAAC 892  
Db |||||  
QY 444 AATGGCTTTGGGTGTGAGAGTGTGGGAAAGAGTACACCATCTGGGCTATAAGA 503  
Db |||||  
QY 893 GTCACTTGGCTTGGATGCCCAACAGTGTGTGACCTCACCTGTAAGGTATGTTTGCAA 952  
Db |||||  
QY 504 GGCACCTGGCTTCCATCGGCCAGCAGTGGGACCTCACCTGTGGGTCTGTGCCCTGG 563  
Db |||||  
QY 953 CTTTGTGAAGCAGCGAGTGTCTTGGAGACCTTAAATCTCATGC---AGSCAAGTCTGT 1009  
Db |||||  
QY 564 AGCTAGGGAGCAGCGAGTGTCTTGGAGACCTTAAATCTCATGC---AGSCAAGTCTGT 623  
Db |||||  
QY 1010 CTGTTGGGTTTAAAGAAAAAGCAACAGTGTGCACTGTGAGAAAGAGTCTTCTCTGTCA 1069  
Db |||||  
QY 624 CTAGCGGAACCAAGGAAAGAGCAGCAGTGTGAGGAGTGTGAGGAGTGTCTTCTACACC 683  
Db |||||  
QY 1070 GAAAGGATGTCCGAGACACATGTTGGTGTGCACTGTGAGAAAGAGTCTTCTCTGTCACT 1129  
Db |||||  
QY 684 GGAAGGATGTGCGACGCGCACCCTGGTGGTCCACACAGGATGCAAGGATCTTCTGTGCCAT 743  
Db |||||  
QY 1130 ATTGTGCAAGAGATTGGGCGAAAGGATCACTGACTGCATATGAGAGAGTCAACA 1189  
Db |||||  
QY 744 TCTGTGCCAGAGATTGGGCGCAAGGATCACTTCCACCGGCATACCAAGAGAGCCACT 803  
Db |||||

QY 1190 ATCAAGAGCTTCTGAAGGTCAAAACAGAACCAAGTGGATTTCCTTGACCCATT 1241  
 Db 804 CACAGGAGCTGATGAAGAGAGCTTGCAGACCGGAGACCTTCTGAGCACCTT 855

RESULT 12  
 ABA52644/c  
 ID ABA52644 standard; DNA; 475 BP.  
 XX  
 AC ABA52644;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Human foetal liver single exon nucleic acid probe #949.  
 XX  
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157277-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00669.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 WPI; 2001-483447/52.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX  
 PS Claim 1; SEQ ID NO 949; 639pp + sequence listing; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC foetal liver. The present sequence is a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 475 BP; 75 A; 129 C; 142 G; 129 T; 0 other;

Query Match  
 Best Local Similarity 3.7%; Score 267; DB 22; Length 475;  
 Matches 352; Conservative 74.1%; Pred. No. 4.3e-41;  
 Mismatches 120; Indels 3; Gaps 1;

QY 733 ACTCATCTCTCTGAGAAACCCACCAAGTGTATTAATTGTGAGAAATGTTTCACCGAAA 792  
 Db 475 ACCATTCTCCCGAAGAAATCCACAGTGTCTCACTGTGAGAGACGTTCAACCGGAAA 416  
 QY 793 GATCATCTGAAGAATCACTCCATACACACACCCCTTAAAGAGAGCGTTTAACTGGGAA 852  
 Db 415 GACCACCTGAAGAAACCACTCCAGACCCACACCCCAAAATGGCTTTGGTGTGAG 356  
 QY 853 GAATGTGGCAGAACTCAATATACCAAGCTTGGATTAAAGCTCACTTGGCCCTTCATGCC 912  
 Db 355 GAGTGTGGGAGAGTACACACCATCTGGGCTATAAGAGGCACCTGGCCCTCCATGCG 296

QY 913 GCACAAGTGGTGACCTCACCTGTAAAGGTATGTTTGCAAACTTTTGAAGCAGCGGAGNG 972  
 Db 295 GCACAGTGGGGACCTCACCTGTGGGTCTGTGCTCCCTGGAGCTAGGAGACCCGAGGTG 236  
 QY 973 CTTCTGGAGCACCTTAAATCTCATGC---AGCAAGTGTCTGGTGGGTTAAAGAAAAA 1029  
 Db 235 CTACTGGACCACTTCAAGCCCATGCGGAAGAGAGCCCTTAGCGAACAAGGAAAG 176  
 QY 1030 AAGCACAGTGCAGACATTGTGATCGCGGTTCTACACCCGAAAGATGTCCGGAGACAC 1089  
 Db 175 AAGCACAGTGCAGACCTGTGAAGATGCTTCTACCCGGAAGATGTGCGACGCCAC 116  
 QY 1090 ATGTTGTGACACTGGAAGAAAGACTTCTCTGTCACTATTGTGACAGAGATTTCGG 1149  
 Db 115 CTGTTGTGTCACAGAGATGCAAGACTTCTCTGTGCGGATTTCTGTGCCAGAGATTGGG 56  
 QY 1150 CGAAGGATCATCTGACTCGACATATGAAGAGAGTCAATCAAGAGCTTCTGA 1204  
 Db 55 CGCAAGGATCACCTCACCCGCGCATACCAAGAGAGCCCACTCACAGGAGCTGATGA 1

RESULT 13  
 ABA22432/c  
 ID ABA22432 standard; DNA; 475 BP.  
 XX  
 AC ABA22432;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Probe #898 for gene expression analysis in human heart cell sample.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system; probe;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00666.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 WPI; 2001-488899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 XX  
 PS Claim 1; SEQ ID No 898; 530pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart. The  
 CC present sequence is one such probe. The probes may be used for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from the human heart via microarrays. By measuring gene expression, the  
 CC probes are useful for predicting, diagnosing, grading, staging,  
 CC monitoring and prognosing diseases of the human heart and vascular system  
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 CC congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO

at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 475 BP; 75 A; 129 C; 142 G; 129 T; 0 other;

Query Match 3.7%; Score 267; DB 22; Length 475;  
Best Local Similarity 74.1%; Pred. No. 4.3e-41;  
Matches 352; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

733 ACTCATCTCTCGAGAAAACCCACCAAGTGTAAATTATCTGAGAAATATGTTTACCGGAAA 792  
475 ACCCATTTCTCCACAGAAATCTCACCAGTGTGCTCACTGTGAGAGACGTTTCAACCGGAAA 416

793 GATCATCTGAAGAATCACCTCCATACACACGACCCCTAACAAAGAGACGTTTAAAGTCGAA 852  
415 GACCACTTGA AAAACCACTTCAGACCCACGACCCCAACAAATGGCTTTGGGTGTGAG 356

853 GAATGTGGCAAGAACTACAAATACCAAGCTTGAATTAAAGTCACTTGGCCTTGCATGCC 912  
355 GAGTGTGGGAAGAAGTACAACACCATGCTGGGCTATAAGAGGCACCTGGGCCCTTCCATGCG 296

913 GCACAAGTGGTGACCTCACCTGTAAGGTATGTTTGCAAACTTTTGAAAGCACGGGAGTG 972  
295 GCAGCAGTGGGGACCTCACCTGTGGGTCTGTGCCCTGGAGCTAGGAGCACCGAGGTG 236

973 CTTCTGGAGCACCTTAAATCTCATGC---AGGCAAGTCTGTGGTGGGTTTAAGAAAAA 1029  
235 CTACTGGACCACTCAAAAGCCCATGGGAAGAAAGCCCTAGCGGAACCAAGGAAAG 176

1030 AAGCACAGTGGCAACATTTGTGATCGCCGTTCTACACCCGAAAGGATGTCCGGAGACAC 1089  
175 AAGCACAGTGGCACCACTGTGAAGATGTTTACACCCGGAAGGATGTGGCAGCCAC 116

1090 ATGTTGTGCACACTGGGAAGAAAGGACTTCTCTGTCTAGTATTGTGCACAGAGATTGGG 1149  
115 CTGTTGTGCACACAGATGCAGGACTTCTCTGTGCCAGTTCTGTGCCACAGAGATTGGG 56

1150 CGAAGGATCACCTGATCGACATATGAAGAAGAGTCAATCAAGAGCTTCTGA 1204  
55 CGCAAGGATCACCTCAGCCGGCATACCAAGAAGACCCACTCACAGGAGCTGATGA 1

RESULT 14  
AAK00905/C  
ID AAK00905 standard; DNA; 475 BP.  
XX AAK00905;  
AC AAK00905;  
XX 05-NOV-2001 (first entry)  
XX Human brain expressed single exon probe SEQ ID NO: 896.  
XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX Homo sapiens.  
OS WO200157275-A2.  
PN 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00667.  
PF 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.

XX 27

```
PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488900/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX
XX PS Example 4; SEQ ID NO: 917; 658pp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX CC the probes of the invention.
XX
XX SQ Sequence 475 BP; 75 A; 129 C; 142 G; 129 T; 0 other;
```

Query Match 3.7%; Score 267; DB 22; Length 475;  
Best Local Similarity 74.1%; Pred. No. 4.3e-41;  
Matches 352; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

```
QY 733 ACTCATTTCTCTGAGAAACCCACAGTGTATTTGTGAGAAAATGTTTCAACCGGAA 792
Db 475 ACCCAATTCCTCCAGAAATCTCACCAGTGTGCTACTGTGAGAGACGTTCAACCGGAA 416
QY 793 GATCATCTGAAGAACTCACTCCATACACAGCCTTACAAAGAGACGTTTAAAGTGGAA 852
Db 415 GACCACTTGAAGAAACCACTCCAGACCCACGACCCCAACAAATGGCCTTTGGGTGTGAG 356
QY 853 GAATGTGGCAAGAACTACAATACCAAGCTTGGATTTAAACGTCCTTGGCCTTGCATGCC 912
Db 355 GAGTGTGGGAAGAGACTACAACCATGCTGGCTATTAAGAGGCACCTGGCCCTCCATGCG 296
QY 913 GCAACAGTGTGACCTCACCTGTAAAGTATGTTTGCAAAACCTTTGAAAGCACGCGGAGTG 972
Db 295 GCCAGCAGTGGGACCTCACCTGTGGGTCTGTGCGCTAGGAGGACCGAGGTG 236
QY 973 CTTCCTGGAGCACCTTAATCTCATGC---AGCAAGTGTGTGGTGGGTTAAAGAAAA 1029
Db 235 CTACTGGACCACTTCAAGCCCATGCGGAAGAGAGCCCTTAGCGGAACCAAGGAAAAAG 176
QY 1030 AAGCACAGTGCAGAACTTGATGCGCGGTTCTACACCCGAAAGGATGCCGGAGACAC 1089
Db 175 AAGCACAGTGCAGACCACTGTGAAGATGCTTCTACACCCGGAAGGATGCGGACGCCAC 116
QY 1090 ATGGTGTGACACCTGGAAGAAAGACTTCTCTGTGATTTGTGCACAGAGATTGGG 1149
Db 115 CTGGTGTGCCACACAGGATGCAAGGACTTCTGTGCCAGTTCTGTGCCAGAGATTGGG 56
QY 1150 CGAAGGATCACCTGACTCGACATATGAGNAGAGTCAATCAAGAGCTTCTGA 1204
Db 55 CGCAGGATCACCTCACCCGGCATACCAAGAAAGACCCACTCACAGGAGCTGATGA 1
```

Search completed: January 31, 2004, 06:29:02  
Job time : 1686 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 06:29:05 ; Search time 13516 Seconds  
(without alignments)  
13150.239 Million cell updates/sec

Title: US-09-242-772-116  
Perfect score: 7313  
Sequence: 1 ggcagcgcatacactacaat.....tatgaataaaatctcgtgcc 7313

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estm.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_hic.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_hic.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pln.\*
- 20: em\_gss\_vrt.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_mam.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_rod.\*
- 26: em\_gss\_phg.\*
- 27: em\_gss\_vri.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	806.4	11.0	1201	13	BX424854	BX424854
2	793.6	10.9	866	13	BX370812	BX370812
3	758	10.4	909	13	BUI181917	BUI181917
4	704.4	9.6	734	14	CA411972	CA411972

5	688	9.4	790	9	AUI138448	AUI138448
6	687	9.4	742	14	CA413047	CA413047
7	685	9.4	693	12	BM724590	BM724590
8	666	9.1	1201	13	BX402039	BX402039
9	662.2	9.1	736	9	AL036679	AL036679
10	629	8.6	907	10	BG613583	BG613583
11	590.8	8.1	616	13	BUE68683	BUE68683
12	570	7.8	570	9	AW339499	AW339499
13	565.4	7.7	922	10	BG253502	BG253502
14	556.4	7.6	659	29	AG152578	AG152578
15	549.8	7.5	673	29	AG120959	AG120959
16	540	7.4	540	9	AA706806	AA706806
17	533.6	7.3	669	14	BY734789	BY734789
18	529.4	7.2	560	9	AI808840	AI808840
19	518	7.1	518	10	BF115791	BF115791
20	514	7.0	818	10	BF243649	BF243649
21	505.4	6.9	597	10	AW977860	AW977860
22	504	6.9	723	29	BZ845384	BZ845384
23	498.4	6.8	500	10	BF507694	BF507694
24	494.2	6.8	719	29	BZ922587	BZ922587
25	482	6.6	482	9	AA707386	AA707386
26	482	6.6	494	9	AW007402	AW007402
27	464	6.3	466	9	AA628501	AA628501
28	462.2	6.3	467	9	AW274762	AW274762
29	448.2	6.1	458	9	AW002999	AW002999
30	439	6.0	449	14	CA771364	CA771364
31	438.2	6.0	454	9	AA418251	AA418251
32	437.6	6.0	608	10	BG378156	BG378156
33	434.2	5.9	683	14	BY719727	BY719727
34	434	5.9	492	28	B94684	B94684
35	430	5.9	432	10	BF432133	BF432133
36	427.4	5.8	460	9	AV745206	AV745206
37	423	5.8	423	9	AI334157	AI334157
38	422.2	5.8	427	9	AI223324	AI223324
39	421.6	5.8	642	13	B0210438	B0210438
40	417.2	5.7	421	14	D59273	D59273
41	406.6	5.6	489	9	AV745506	AV745506
42	400.8	5.5	481	10	BF710381	BF710381
43	391.2	5.3	4767	11	AK040626	AK040626
44	388	5.3	522	12	BI298426	BI298426
45	373	5.1	373	9	AA418372	AA418372

## ALIGNMENTS

RESULT 1	BX424854	1201 bp	mRNA	linear	EST 13-MAY-2003
LOCUS	BX424854	Homo sapiens	NEUROBLASTOMA	Homo sapiens	cdna clone
DEFINITION	XCL0BB001ZD02	5-PRIME, mRNA	sequence.		
ACCESSION	BX424854				
VERSION	BX424854.1	GI:30647825			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1201)				
AUTHORS	Li, W.B., Gruber, C., Jesse, J., and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished				
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4473.r For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=XCL0BB001ZD02RP1&cluster=4473.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600				



1493 CTACCTCATATGCAATTTCTATTCTGAAAGAACACAGCCATTAAGGGGAAATGAGA 1552  
1088 CTACCTCATATGCAATTTCTATTCTGAAAGAACACAGCCATTAAGGGGAAATK--AG 1145  
1553 GTTACCTGATGAGTTACAAGTGGCGTCCCTCTTCTATCCCAAGATTTCTCAAGCATC 1610  
1146 RKTCCGGATGAGTACAGKGGC-TKCCCTCTCATCC-ARATWCTCAAGCATS 1201  
RESULT 2  
BX370812 866 bp mRNA linear EST 08-MAY-2003  
BX370812 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DL012YM24 5-PRIME, mRNA sequence.  
ACCESSION BX370812  
VERSION BX370812.1 GI:30445558  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 866)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 4473.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAG062D02\_CS05868\_1&cluster=4473.r.  
Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paraday Avenue Genoscope sequence ID : CS0BAG062D02\_CS05868\_1.  
FEATURES  
Location/Qualifiers  
1..866  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL012YM24"  
/cell\_type="B CELLS (RAMOS CELL LINE)"  
/clone\_lib="RAMOS CELL LINE"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 235 a 218 c 177 g 236 t  
ORIGIN  
Query Match 10.9%; Score 793.6; DB 13; Length 866;  
Best Local Similarity 98.5%; Pred. No. 3.1e-72; Indels 4; Gaps 4;  
Matches 843; Conservative 0; Mismatches 9; Indels 4; Gaps 4;  
984 CCTTAATCTCATGACGCAAGTCTCTGTGGGGTTAAAGAAACCAACGACGTCGA 1043  
14 CCTTAATCTCATGCA-GCAAGTCTGTGGGGTTAAAGAAACCAACGACGTCGA 71  
1044 ACATTGTGATCGCGGTTCTACACCGAAGATGTCCGAGACATGTTGGTGCACAC 1103  
72 ACATTGTGATCGCGGTTCTACACCG-AAGATGTCCGAGACATGTTGGTGCACAC 130  
1104 TGAAGAAAGGACTTCTCTGTGATGTCACAGATTTGGCGGAAAGGATCACCT 1163  
131 TGAAGAAAGGACTTCTCTGTGATGTCACAGATTTGGCGGAAAGGATCACCT 190  
1164 GACTCCACATATCAAGAAAGTCAATCAAGAGCTTCTGAAGTCAAAACAGAACAGT 1223  
191 GACTCCACATATCAAGAAAGTCAATCAAGAGCTTCTGAAGTCAAAACAGAACAGT 250

Paraday Avenue Genoscope sequence ID : XCLOBB0012D02RP1.

FEATURES  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="XCLOBB0012D02"  
/tissue\_type="NEUROBLASTOMA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."  
BASE COUNT 354 a 284 c 274 g 271 t 18 others  
ORIGIN  
Query Match 11.0%; Score 806.4; DB 13; Length 1201;  
Best Local Similarity 97.0%; Pred. No. 1.3e-73; Indels 7; Gaps 6;  
Matches 871; Conservative 12; Mismatches 8; Indels 7; Gaps 6;  
714 ATTCAAGGCACATGCTACTCTCTCTGAGAAACCCCAAGTGAATTTATTGTA 773  
310 AGTAATCAGGCATGCTACTCTCTCTGAGAAACCCCAAGTGAATTTATTGTA 369  
774 GAAATGTTTCACCGAAGATCATCTGAAGATCACTCATACACGACCTTACAA 833  
370 GAAATGTTTCACCGAAGATCATCTGAAGATCACTCATACACGACCTTACAA 429  
834 AGAGAGCTTTAAGTGGCAAGATGTGCAAGAACTACAATCAAGCTTGGATTAAACG 893  
430 AGAGAGCTTTAAGTGGCAAGATGTGCAAGAACTACAATCAAGCTTGGATTAAACG 489  
894 TCATTGGCTTGCATCCGCAAGTGTGACCTCACCTGTAAAGTATGTTGCAAC 953  
490 TCATTGGCTTGCATCCGCAAGTGTGACCTCACCTGTAA-GTATGTTGCAAC 548  
954 TTTTGAAGCAGCGGAGTCTCTTGGAGCCTTTAAATCTCATGCGGCAAGTCTGTG 1013  
549 TTTTGAAGCAGCGGAGT-CTTCTGGAGCCTTTAAATCTCATGCGGCAAGTCTGTG 607  
1014 TGGGGTTAAAGAAAGACAGTGTGCAATGTGATCGCGGTTCTACACCGAAA 1073  
608 TGGGGTTAAAGAAAGACAGTGTGCAATGTGATCGCGGTTCTACACCGAAA 667  
1074 GGATGTCCGGAGACATGTTGTCACACTGGAAGAAAGACTTCTCTGTGATTTG 1133  
668 GGATGTCCGGAGACATGTTGTCACACTGGAAGAAAGACTTCTCTGTGATTTG 727  
1134 TGCACAGATTTGGCGGAAAGGATCACCTGACTCGAATATGAAGAGAGTCAATCA 1193  
728 TGCACAGATTTGGCGGAAAGGATCACCTGACTCGAATATGAAGAGAGTCAATCA 787  
1194 AGAGCTTCTGAAGTCAAAACAGACAGTGTGATTTCTTGACCCATTTACCTGCAATGT 1253  
788 AGAGCTTCTGAAGTCAAAACAGACAGTGTGATTTCTTGACCCATTTACCTGCAATGT 847  
1254 GTCTGTGCTATAAAGAGAGCTCTCTCCGCTGATGTCTTACCTTCCAGTGAATGTT 1313  
848 GTCTGTGCTATAAAGAGAGCTCTCTCCGCTGATGTCTTACCTTCCAGTGAATGTT 907  
1314 ATCAAGCCATTACAAACACTTTGAGTTAAACCTCTCAACACTCCATTTCAGTCCAT 1373  
908 ATCAAGCCATTACAAACACTTTGAGTTAAACCTCTCAACACTCCATTTCAGTCCAT 967  
1374 GCAGAGCTGGGATCTGCCCAACAATGATCAACTTTACCTTTGGGATGACATGCC 1433  
968 GCAGAGCTGGGATCTGCCCAACAATGATCAACTTTACCTTTGGGATGACATGCC 1027  
1434 AATAGATGAGACACTGTTTCATCC-TCTCACCACCTTTCTTTCAAATATCCGTTCA 1492  
1028 AATAGATGAGACACTGTTTCATCC-TCTCACCACCTTTCTTTCAAATATCCGTTCA 1087

```

QY 1224 GGAATTCCTGACCAATTTACTGCAATGTGTCTGTGCTATATAAGACGAGCTCTTCC 1283
Db |||||||
QY 251 GGAATTCCTGACCAATTTACTGCAATGTGTCTGTGCTATATAAGACGAGCTCTTCC 310
Db |||||||
QY 1284 GGTGATGTCCTTACCTCCAGTGAACCTGTTATCAAGCCATTCACAAACATTTGCGATT 1343
Db |||||||
QY 311 GGTGATGTCCTTACCTCCAGTGAACCTGTTATCAAGCCATTCACAAACATTTGCGATT 370
Db |||||||
QY 1344 AAACCTCTACAACTCCATTTAGTCCATCGACGAGCTCGGATCTGCCACCAAAATGAT 1403
Db |||||||
QY 371 AAACCTCTACAACTCCATTTAGTCCATCGACGAGCTCGGATCTGCCACCAAAATGAT 430
Db |||||||
QY 1404 CACAACCTTACCTTTGGGAATGACATGCCCAATAGATATGGACACTGTTTCATCCCTCTCA 1463
Db |||||||
QY 431 CACAACCTTACCTTTGGGAATGACATGCCCAATAGATATGGACACTGTTTCATCCCTCTCA 490
Db |||||||
QY 1464 CCACCTTCTCTTTCAAAATATCGGTTTCAGTCTTACCTCTCATATGCAATTTCTATTCCTGAAA 1523
Db |||||||
QY 491 CCACCTTCTCTTTCAAAATATCGGTTTCAGTCTTACCTCTCATATGCAATTTCTATTCCTGAAA 550
Db |||||||
QY 1524 AGACACGCAATTAAGGGGAAATTTGAGAGTTACTGATGGAGTTACAAGTGGCGTGC 1583
Db |||||||
QY 551 AGACACGCAATTAAGGGGAAATTTGAGAGTTACTGATGGAGTTACAAGTGGCGTGC 610
Db |||||||
QY 1584 CTCTTCATCCCAAGATCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCTCAGAT 1643
Db |||||||
QY 611 CTCTTCATCCCAAGATCTCAAGCATCGTCATCATCTAAGCTAGGTTGGATCCTCAGAT 670
Db |||||||
QY 1644 TGGGTCCTTAGATGATGGTGCAGGAGCTCTCCCTATCCAAAGCTCTATCTCCATCAG 1703
Db |||||||
QY 671 TGGGTCCTTAGATGATGGTGCAGGAGCTCTCCCTATCCAAAGCTCTATCTCCATCAG 730
Db |||||||
QY 1704 TGACCCCTTAAACACACAGCATTTGATTTTCTCAGTTGTTTAAATTTTCATACCTTTAA 1763
Db |||||||
QY 731 TGACCCCTTAAACACACAGCATTTGATTTTCTCAGTTGTTTAAATTTTCATACCTTTAA 790
Db |||||||
QY 1764 TGGTCTCTCTATATCTCTATCATAGTGGGAGCTTGAATGAGCTATTCAGGAGCA 1823
Db |||||||
QY 791 TGGTCTCTCTATATCTCTATCATAGTGGGAGCTTGAATGAGCTATTCAGGAGCA 849
Db |||||||
QY 1824 AGCATTCTCTCTGTT 1839
Db |||||||
QY 850 AGCATTCTCTGTT 865
Db |||||||

RESULT 3
BUI181917
LOCUS
DEFINITION BUI181917 909 bp mRNA linear EST 04-SEP-2002
5' mRNA sequence.
ACCESSION BUI181917
VERSION BUI181917.1 GI:22695901
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 909)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13451 row: o column: 20
High quality sequence stop: 683.

```

## FEATURES

```

source
Location/Qualifiers
1..909
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:6137707"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 329 a 132 c 142 g 305 t 1 others
ORIGIN
Query Match 10.4%; Score 758; DB 13; Length 909;
Best Local Similarity 95.8%; Pred. No. 1.3e-68;
Matches 844; Conservative 0; Mismatches 26; Indels 11; Gaps 6;
QY 3255 ATTAGAAAACCTACCTGATGAACCTACCAAAAGTTTAAAGTGACTAGAAATATATACAGTA 3314
Db |||||||
QY 3315 AATCCCACTTTCATGCACTCTCGGAAATGATAGGAGTATTCGAAATAGTTGAGTTTG 3374
Db |||||||
QY 61 AATCCCACTTTCATGCACTCTCGGAAATGATAGGAGTATTCGAAATAGTTGAGTTTG 120
Db |||||||
QY 3375 TAGAGGGTAAACAAAGTAAAGTAAACAAACCTATCTTGGTTAAACATCAAAATTAACATTG 3434
Db |||||||
QY 121 TAGAGGGTAAACAAAGTAAAGTAAACAAACCTATCTTGGTTAAACATCAAAATTAACATTG 180
Db |||||||
QY 3435 AGAATATATTTATTCACCTGAATAATTTATAGGCTTTTCTCAGATTTAGCAATTAACCTAG 3494
Db |||||||
QY 181 AGAATATATTTATTCACCTGAATAATTTATAGGCTTTTCTCAGATTTAGCAATTAACCTAG 240
Db |||||||
QY 3495 ATCTTCTTAAAGGTCTAATTAATATATTTTCTAAGGGTCAGTTGGGACATTAACCTAG 3554
Db |||||||
QY 241 ATCTTCTTAAAGGTCTAATTAATATATTTTCTAAGGGTCAGTTGGGACATTAACCTAG 300
Db |||||||
QY 3555 AAACATATCTATTAGCACTTGTAAACCTTATTTAGGACCTTTTCGGTTGGGATGG 3614
Db |||||||
QY 301 AAACATATCTATTAGCACTTGTAAACCTTATTTAGGACCTTTTCGGTTGGGATGG 360
Db |||||||
QY 3615 GGGCAAGGGTGGGAGTTTGTAGAGAGTATATATCTCTTTTAAAAAAGGAGGAGGAG 3673
Db |||||||
QY 361 GGGCAAGGGTGGGAGTTTGTAGAGAGTATATATCTCTTTTAAAAAAGGAGGAGGAG 420
Db |||||||
QY 3674 AAATATTTCTGAGCACTCATTTAGCCCTATATGAAACTTCTTCTCTTTTCTAGGCGAG 3733
Db |||||||
QY 421 AAATATTTCTGAGCACTCATTTAGCCCTATATGAAACTTCTTCTCTTTTCTAGGCGAG 480
Db |||||||
QY 3734 TTATCACTGAGATTTGCAATGTTTACCAAGATTTTCAAAATAGTGCAGATTACTGAA 3793
Db |||||||
QY 481 TTATCACTGAGATTTGCAATGTTTACCAAGATTTTCAAAATAGTGCAGATTACTGAA 540
Db |||||||
QY 3794 TATAATACATTTTAAATATTTTGGGAGTAGTATATTTTGTGAGAAATGTAATTTGTA 3853
Db |||||||
QY 541 TATAATACATTTTAAATATTTTGGGAGTAGTATATTTTGTGAGAAATGTAATTTGTA 600
Db |||||||
QY 3854 ATATGTAAATGGGGCTTCAATATATATATATATATATATATATATATATATATATATAT 3913
Db |||||||
QY 601 ATATGTAAAT--GGGGCTTCAATATATATATATATATATATATATATATATATATATATAT 659
Db |||||||
QY 3914 CATACCGCACTTCATAGATCAAAAGTTGCTCTGAGGAGCTTTGGCTCTGATATTTT 3973
Db |||||||
QY 660 CATACCGCACTTCATAGATCAAAAGTTGCTCTGAGGAGCTTTGGCTCTGATATTTT 719
Db |||||||
QY 3974 ATCATGCTCTATTTTAAATATTTTAACTCTGAGCAGTAG--TTTTTATCTATGATTTAA 4032
Db |||||||
QY 720 ATCATGCTCTATTTTAAATATTTTAACTCTGAGCAGTAGTTTATATATCTATGATTTAA 779
Db |||||||
QY 4033 A--TTTTATATGAAAAATTAACATTTATTAAAAA---GTGTCTTCCAAAGGCAATAAAA 4087
Db |||||||

```

Db 780 AATTTTATATGAAAAAATTACTTTTATTTAAAAAGTGTGTTTCAAAAGCCATTAAAA 839

Qy 4088 ---TTATATATCTTAAATAGGAAGTACATTTTTTAAATTTTT 4125

Db 840 ATTATATATCTTAAATAGGAAGTACCTTTTTTAAATTTTT 880

RESULT 4

CA411972/c

LOCUS CA411972 734 bp mRNA linear EST 07-NOV-2002

DEFINITION UI-H-EZO-bal-m-01-0-UI.s1 NCI CGAP Ch1 Homo sapiens cDNA clone

UI-H-EZO-bal-m-01-0-UI 3', mRNA sequence.

ACCESSION CA411972

VERSION CA411972.1 GI:24774623

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 734)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-xemail.nih.gov

Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of Orthopedics

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu

The following repetitive elements were found in this cDNA sequence: 1-22, >AT rich#Low\_complexity (matched complement)

Seq primer: M13 FORWARD

POLYA=Yes.

FEATURES

source Location/Qualifiers

1..734

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="UI-H-EZO-bal-m-01-0-UI"

/tissue\_type="Chondrosarcoma Grade II"

/dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="NCI CGAP Ch1"

/note="Organ: Left Pelvis; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP Ch1 is a cDNA library containing the following tissue(s): Chondrosarcoma Grade II. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dr primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (AT)18 tail. The sequence tag for this library is TGATCATCGCT.

TAG LIB=UI-H-EZO

TAG TISSUE=grade-2-chondrosarcoma

TAG SEQ=ATCTAATATG"

BASE COUNT 216 a 212 c 179 g 214 t 2 others

ORIGIN

Query Match 9.6%; Score 704.4; DB 14; Length 734;

Best local similarity 99.6%; Pred. No. 4.4e-63;

Matches 716; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1424 TGACATGCCAATAGATATGGACACTGTTCATCCCTCTCACACCTTTCTTTCAAAATATC 1483



DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 593  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-E01-aiz-j-24-0-UI"  
/tissue\_type="fetal eye"  
/dev\_stage="fetal"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-E01"  
/notes="Organ: eye; Vector: p77T3-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-E-E01 is a normalized cDNA library containing the  
following tissue(s): fetal eye. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into p77T3-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dfr)18 tail. The sequence tag for this library is  
CGCGTATACC. This library was created for the program: Gene  
Discovery in the Visual System, supported by National Eye  
Institute (NEI)."  
BASE COUNT 208 a 101 c 112 g 272 t  
ORIGIN  
Query Match 9.4%; Score 685; DB 12; Length 693;  
Best Local Similarity 100.0%; Pred. No. 4.4e-61;  
Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5133 TGAGTGTGAAGAGTCTTGGCACTGTGAAGCAATGCTATCCATGATGATATACAGTACCT 5192  
Db 9 TGAGTGTGAAGAGTCTTGGCACTGTGAAGCAATGCTATCCATGATGATATACAGTACCT 68  
QY 5193 TTATAGTTATCGATCAGTCTTAAACATTTTCAATTTTAAATCTTATTTACCAAGTTCAGTTT 5252  
Db 69 TTATAGTTATCGATCAGTCTTAAACATTTTCAATTTTAAATCTTATTTACCAAGTTCAGTTT 128  
QY 5253 TTTAAACCTTCAATTTGCTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5312  
Db 129 TTTAAACCTTCAATTTGCTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 188  
QY 5313 ATTTAGTGTCTTTTCAAGCTGTGATTTTTCCTTATTTTTCCTTATTTTTCCTTATTTTTC 5372  
Db 189 ATTTAGTGTCTTTTCAAGCTGTGATTTTTCCTTATTTTTCCTTATTTTTCCTTATTTTTC 248  
QY 5373 TTAGTCATTTTGGGAATATAGTATATTTGTTTAAATTTGGACAGTACGGGTTTTTAA 5432  
Db 249 TTAGTCATTTTGGGAATATAGTATATTTGTTTAAATTTGGACAGTACGGGTTTTTAA 308  
QY 5433 AACCATATATGACTGAAACATGAGCAGCGGATGCTTTTATTAAGCTAATAAGTAAAT 5492  
Db 309 AACCATATATGACTGAAACATGAGCAGCGGATGCTTTTATTAAGCTAATAAGTAAAT 368  
QY 5493 GTTAAAGAGTACATATTTTTCAGGATCGTTTCATCTAGTGAGCAATACACATATATAGGCC 5552  
Db 369 GTTAAAGAGTACATATTTTTCAGGATCGTTTCATCTAGTGAGCAATACACATATATAGGCC 428  
QY 5553 AATATTTTAAATAAGTACGCTTGGTCAACCTCTATATACATATATTAAGTATATA 5612  
Db 429 AATATTTTAAATAAGTACGCTTGGTCAACCTCTATATACATATATTAAGTATATA 488  
QY 5613 GCATTTTCAAAATGAATCTAAACCTTTTACAGAACTTTCTATAGGTTATGCTTTTATTT 5672  
Db 489 GCATTTTCAAAATGAATCTAAACCTTTTACAGAACTTTCTATAGGTTATGCTTTTATTT 548

FEATURES  
source  
683 AGCTTTTCCACNAAGTGACTGAACCTAGATGGAAGTGGCTTTATATAGGTG 624  
QY 6715 TCGAGTGAAGACATGCTGTATTTACTAAACAGCCCTTTGAATTTAAACAAAACCTGGGAA 6774  
Db 623 TCGAGTGAAGACATGCTGTATTTACTAAACAGCCCTTTGAATTTAAACAAAACCTGGGAA 564  
QY 6775 TCCATTAGGAACCGGATTCATCATACCTGACATGAAGCTGAGCTGCTGAATTTGTTATTT 6834  
Db 563 TCCATTAGGAACCGGATTCATCATACCTGACATGAAGCTGAGCTGCTGAATTTGTTATTT 504  
QY 6835 TTAGCTAATGAAAGAGTGTGGACTAGTACTCTAAATGTTCTAATGATAAAGTTTGG 6894  
Db 503 TTAGCTAATGAAAGAGTGTGGACTAGTACTCTAATGATAAAGTTTGG 444  
QY 6895 AGTCAAAATAGAAAAGAAAATCTGCAATTCAGGCGCGAATTTGTATATTTTATTCG 6954  
Db 443 AGTCAAAATAGAAAAGAAAATCTGCAATTCAGGCGCGAATTTGTATATTTTATTCG 384  
QY 6955 ATTAAATATGCTATTTCTGTAATTTGGGAATCAAGTGGCTTATCATGATATATCGTGA 7014  
Db 383 ATTAAATATGCTATTTCTGTAATTTGGGAATCAAGTGGCTTATCATGATATATCGTGA 324  
QY 7015 CTTAAATATGCTATTTCTGTAATTTGGGAATCAAGTGGCTTATCATGATATATCGTGA 7074  
Db 323 CTTAAATATGCTATTTCTGTAATTTGGGAATCAAGTGGCTTATCATGATATATCGTGA 264  
QY 7075 TTTTGTGTGTATAGCTCTGTAATTTGGGAATCAAGTGGCTTATCATGATATATCGTGA 7134  
Db 263 TTTTGTGTGTATAGCTCTGTAATTTGGGAATCAAGTGGCTTATCATGATATATCGTGA 204  
QY 7135 CATTTTAAACATTCACATCCAAAGAGCAGTATTTTATTTTCCATATACCAAGATTTAAA 7194  
Db 203 CATTTTAAACATTCACATCCAAAGAGCAGTATTTTATTTTCCATATACCAAGATTTAAA 144  
QY 7195 ATATTAATTTCTGCTAATTAATAATAGTACTGAGCTTCTTGTGGCTACAGTGTAT 7254  
Db 143 ATATTAATTTCTGCTAATTAATAATAGTACTGAGCTTCTTGTGGCTACAGTGTAT 84  
QY 7255 GTTTGCTCTAAGATAAGATATGTGAATTCACAAAATATATGAATAAAT 7305  
Db 83 GTTTGCTCTAAGATAAGATATGTGAATTCACAAAATATATGAATAAAT 33  
RESULT 7  
BM724590 693 bp mRNA linear EST 01-MAR-2002  
LOCUS UI-E-E01-aiz-j-24-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone  
DEFINITION UI-E-E01-aiz-j-24-0-UI 5', mRNA sequence.  
ACCESSION BM724590  
VERSION BM724590.1 GI:19045921  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 693)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PUBMED 889548  
COMMENT Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa

QY 5673 TTAAGACTTATTATTAATCAAGTGCCATTAGATGATATATATAGGCTTTTGATATATA 5732  
 Db 549 TTAAGACTTATTATTAATCAAGTGCCATTAGATGATATATATAGGCTTTTGATATATA 608  
 QY 5733 ATGCTTTGTGTACAAAATGGTAGATGTTTAAACAGGTACATTTTACAGTGTCTT 5792  
 Db 609 ATGCTTTGTGTACAAAATGGTAGATGTTTAAACAGGTACATTTTACAGTGTCTT 668  
 QY 5793 CTTATCAATTTGCTATATTGTCACAG 5817  
 Db 669 CTTATCAATTTGCTATATTGTCACAG 693

## RESULT 8

BX402039

## LOCUS

DEFINITION BX402039 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 Homo sapiens cDNA clone CSODL012YM24 5-PRIME, mRNA sequence.

## ACCESSION

BX402039

## VERSION

BX402039.1

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## AUTHORS

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

## TITLE

Full-length cDNA libraries and normalization

## JOURNAL

Unpublished

## COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

4473.r For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODL012BM12QPI&amp;cluster=4473.r. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CSODL012BM12QPI.

## FEATURES

source

1..1201

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CSODL012YM24"

/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"

/cell\_line="RAMOS CELL LINE"

/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT

25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo (dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 319 a 260 c 297 g 273 t 52 others

## ORIGIN

Query Match

Best Local Similarity 9.1%; Score 666; DB 13; Length 1201;

Matches 918; Conservative 6; Mismatches 6; Indels 212; Gaps 4;

QY 1 GGCAGCGATACATCAATGCTCTCTGAAAGAGCGCTAAGGAACAATTCAGAGCCC 60

Db 94 GGCAGCGATACATCAATGCTCTCTGAAAGAGCGCTAAGGAACAATTCAGAGCCC 153

QY 61 GCCGCTCCAGCCGAATATGAGAAAAAATATTAGAAAAATTCGCGGGCGGTGTAG 120

Db 154 GCCGCTCCAGCCGAATATGAGAAAAAATATTAGAAAAATTCGCGGGCGGTGTAG 213

QY 121 GCGGCGACGGCGCGAGGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGTGCTT 180

Db 214 GCGGCGACGGCGCGAGGAGGATGTTAAAGCCCGCGGTGCTCTTGGTGTGCTT 273

QY 181 GCGCGTATTGGCACCAGATGCTTCTTCTGTGACGGTCTATTAAATAGGTTGCTTG 240  
 Db 274 GCGCGTATTGGCACCAGATGCTTCTTCTGTGACGGTCTATTAAATAGGTTGCTTG 333  
 QY 241 CTAGAGTTTGGAGAGGGGCTCAGATTGGCCAAAATGGGAAGATTGGAATCCACTCTCT 300  
 Db 334 CTAGAGTTTGGAGAGGGGCTCAGATTGGCCAAAATGGGAAGATTGGAATCCACTCTCT 393  
 QY 301 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAAAGTCTGAGGCTTTTCCATCAGTAAT 360  
 Db 394 TCCACGAAGAGTCAATGGGACTGGCTAAGATCAAAAGTCTGAGGCTTTTCCATCAGTAAT 453  
 QY 361 CAGTCCCTTTTGTCTTTTACGACCACATGAACCTTGAGAGCCACCTTAAGCTATA 420  
 Db 454 CAGTCCCTTTTGTCTTTTACGACCACATGAACCTTGAGAGCCACCTTAAGCTATA 513  
 QY 421 TCATTAGTGGAGTTGGGAGTCCCAAGTGTCCAAAGAGAGGCTCGTTTAGGCTCG 480  
 Db 514 TCATTAGTGGAGTTGGGAGTCCCAAGTGTCCAAAGAGAGGCTCGTTTAGGCTCG 573  
 QY 481 ATGCCACTGTCTCTCTGTTGATTTGTCAGAGTAAGAGATACCCAGAAAGTCCCTTCA 540  
 Db 574 ATGCCACTGTCTCTCTGTTGATTTGTCAGAGTAAGAGATACCCAGAAAGTCCCTTCA 606  
 QY 541 GGGAAACGTAAAGCGTGGTGAACCAACCAAGAAACCTTCTCTGCCAACTGTGTGAC 600  
 Db 607 ----- 606  
 QY 601 AAGSCCTTTAACAGTGTGAGAAATTAAAGTTTCACTCTCTCACAGGAGAGAGG 660  
 Db 607 ----- 606  
 QY 661 CCCTACAAGTGCATACCAAGACTGCACCAAGCCTTTGTTTCTTAAGTACAAATTACAA 720  
 Db 607 ----- 606  
 QY 721 AGGCACATGCTACTCTCTCTGAGAAACCCCAAGGTGTAATTATTGTGAGAAATG 780  
 Db 607 --GCACATGGCTACTCTCTCTGAGAAACCCCAAGGTGTAATTATTGTGAGAAATG 664  
 QY 781 TTTACCGGAAGATCATCTGAGAAATCACTCCATACACAGCCCTTAACAAAGAGAGC 840  
 Db 665 TTTACCGGAAGATCATCTGAGAAATCACTCCATACACAGCCCTTAACAAAGAGAGC 724  
 QY 841 TTTAAGTGGGAAGATGTGCGAAGACTACATACCAAGCTTGGATTAAACGTCACCTTG 900  
 Db 725 TTTAAGTGGGAAGATGTGCGAAGACTACATACCAAGCTTGGATTAAACGTCACCTTG 784  
 QY 901 GCCTTGCATGCGCAACCAAGTGGTGCACCTCACTCTGTAAGTATGTTGCAAACTTTGAA 960  
 Db 785 GCCTTGCATGCGCAACCAAGTGGTGCACCTCACTCTGTAAGTATGTTGCAAACTTTGAA 844  
 QY 961 AGCAGGAGTGTCTTCT--GGAGCACCTTAAATCTCATGAGCAAGTCTCTGTTGGGT 1019  
 Db 845 AGCAGGAGTGTCTTCTGAGGACACCTTAAATCTCATGAGCAAGTCTCTGTTGGGT 904  
 QY 1020 TAAAGAAAAAGACCAAGTGGGAAACATGTGATGCGCGTTCTACACCCGAAAGATGT 1079  
 Db 905 TAAAGAAAAAGACCAAGTGGGAAACATGTGATGCGCGTTCTACACCCGAAAGATGT 963  
 QY 1080 CCGAGAGACATGTGTGTGACACTGGAAGAAAGAGCTTCTCTGTGATGTTGTGACA 1139  
 Db 964 CCGAGAGACATGTGTGTGACACTGGAAGAAAGAGCTTCTCTGTGATGTTGTGACA 1022  
 QY 1140 GA 1141  
 Db 1023 GA 1024

RESULT 9

AL036879

LOCUS

AL036879

736 bp

mRNA

linear

EST 29-FEB-2000

DEFINITION	DKFZp564P1863_r1 564 (synonym: hfb2) Homo sapiens cDNA clone
ACCESSION	DKFZp564P1863 5', mRNA sequence.
VERSION	AL036879
KEYWORDS	EST.
SOURCE	AL036879.3 GI:5927934
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 736)
JOURNAL	Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
COMMENT	EST (Duesterhoeft, et al.) Unpublished On Jul 7, 1999 this sequence version replaced gi:5866332. Contact: Duesterhoeft A MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Olgan (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available. This clone (DKFZp564P1863) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES	Location/Qualifiers
source	1..736
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="DKFZp564P1863"
	/tissue_type="brain"
	/dev_stage="fetal"
	/lab_host="xl-2blue"
	/clone_lib="564 (synonym: hfb2)"
BASE COUNT	/note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"
ORIGIN	223 a 174 c 157 g 182 t
Query Match	9.1%; Score 662.2; DB 9; Length 736;
Best Local Similarity	98.1%; Pred. No. 9e-59;
Matches 723; Conservative	0; Mismatches 8; Indels 6; Gaps 5;
QY	759 GTGTAATTATGCGAATAATGTTTACCGGAAAGATCATCTGAAGATCACTCCATAC 818
Db	2 GGGTAATTATGCGAATAATGTTTACCGGAAAGATCATCTGAAGATCACTCCATAC 61
QY	819 ACACGACCTTAAACAGACGCTTTAAGTGGCAAGATGTCGCAAGCACTCAATACCAA 878
Db	62 ACACGACCTTAAACAGACGCTTTAAGTGGCAAGATGTCGCAAGCACTCAATACCAA 121
QY	879 GCTTGATTTAAACGTCACCTGCGCTTGCATCGCGCAACAAAGTGTGACCTCACCTGTAA 938
Db	122 GCTTGATTTAAACGTCACCTGCGCTTGCATCGCGCAACAAAGTGTGACCTCACCTGTAA 181
QY	939 GGTATGCTTGCACAACTTTTGAAAGCAGCGGAGTCTTCTGGAGCAGCTTAAATCTCATGC 998
Db	182 GGTATGCTTGCACAACTTTTGAAAGCAGCGGAGTCTTCTGGAGCAGCTTAAATCTCATGC 241
QY	999 AGCAAGTCTGCTGGTGGGTTTAAAGAAAAGCAGCAGTGGCAATGTGATCGCG 1058
Db	242 AGCAAGTCTGCTGGTGGTGTAAAGAAAAGCAGCAGTGGCAATGTGATCGCG 301
QY	1059 GTTCTACACCGAAGATGTCGGAGACACATGGTGTGCACACTGGAAGAAAGACTT 1118
Db	302 GTTCTACACCGAAGATGTCGGAGACACATGGTGTGCACACTGGAAGAAAGACTT 361
QY	1119 CCTCTGTCAGTATTGTGCACAGATTTGGCGGAAAGGATCACTGATGACATATGAA 1178
Db	362 CCTCTGTCAGTATTGTGCACAGATTTGGCGGAAAGGATCACTGATGACATATGAA 421
QY	1179 GAAGAGTCACATCAAGAGCTTCTGAAGCTCAAAACAGAACCAAGTGGATTCTCTTGACCC 1238
Db	422 GAAGAGTCACATCAAGAGCTTCTGAAGCTCAAAACAGAACCAAGTGGATTCTCTTGACCC 481
QY	1239 ATTTACCTGCAATGTGTCTGTGCTCTATATAAAGACGAGCTCTTCCGGTGTATGCTTACC 1298
Db	482 ATTTACCTGCAATGTGTCTGTGCTCTATATAAAGACGAGCTCTTCCGGTGTATGCTTACC 541
QY	1299 TTCCAGTG-AACTGTTATCAAGAGCCATTCAAAACACTTTGAGTTAAACCTCTACAACA 1357
Db	542 TTCCAGTGAACTGTTATCAAGAGCCATTCAAAACACTTTGAGTTAAACCTCTACAACA 601
QY	1358 CTCATTTCAGTCCATGCG-AGAGCTCGGATCTGCCAC-AAATGATCAACAACCTTTACC 1415
Db	602 CTCATTTCAGTCCATGCGAGAGCTCGGATCTGCCAC-AAATGATCAACAACCTTTACC 661
QY	1416 TTTGGGAATGACATGCCCAATAGATAT-GGACACTGTTTCATCCCTCTCACACCTTTCTT 1474
Db	662 TTTGGGAATGACATGCCCAATAGATATGAGACACTGTTCAATCTCTCAACAACCTTTCTT 721
QY	1475 TCAAAATATCGTTTCAGT 1491
Db	722 --CAATATCGTTTCAGT 736
RESULT 10	
BG613583	602641935P1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4772744 5',
LOCUS	mRNA sequence.
DEFINITION	BG613583
ACCESSION	BG613583.1 GI:13664954
VERSION	EST.
KEYWORDS	Homo sapiens (human)
SOURCE	Homo sapiens
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 907)
TITLE	NIH-MGC http://mgt.nci.nih.gov/.
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: CLONETECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM1643 row: f column: 09 High quality sequence stop: 645.
FEATURES	Location/Qualifiers
source	1..907
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="IMAGE:4772744"
	/tissue_type="embryonal carcinoma"
	/lab_host="DH10B (T1 phage-resistant)"
	/clone_lib="NIH_MGC 61"
	/note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccattatggc); Site_2: SfiI (ggccattatggc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGGCACATG-dt(30)BN-3' (Where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."



RESULT 11	DEFINITION
BU686683/C	
LOCUS	

was made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Ronaldo. "

BASE COUNT	180 a	102 c	96 g	192 t	
Query Match	7.8%;	Score 570;	DB 9;	Length 570;	
Best Local Similarity	100.0%;	Pred. No. 2.7e-49;			
Matches 570; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	6339	TCATTATATAAGCTTCAGTGGTACAGATGAACACAGAAATGAATGTTTATCTTCTCGAAAC	6398		
Db	570	TCATTATATAAGCTTCAGTGGTACAGATGAACACAGAAATGAATGTTTATCTTCTCGAAAC	511		
QY	6399	ACTCCTTCAATATATATTTGGATCATGCTGCTAAATGTAAGTCTGGGTACAACTCTTTCATG	6458		
Db	510	ACTCCTTCAATATATATTTGGATCATGCTGCTAAATGTAAGTCTGGGTACAACTCTTTCATG	451		
QY	6459	GTGCTTCAAAACCTTCTCTGCTCATTCAGTCGTATTTTTTATCCATAGAAAAAGGACTAC	6518		
Db	450	GTGCTTCAAAACCTTCTCTGCTCATTCAGTCGTATTTTTTATCCATAGAAAAAGGACTAC	391		
QY	6519	ATTAGGTGTAAGTGTACAATATATTTTATCTGTGACTTAATTTGTCTATTAAACAAC	6578		
Db	390	ATTAGGTGTAAGTGTACAATATATTTTATCTGTGACTTAATTTGTCTATTAAACAAC	331		
QY	6579	TTTTACACCAACCAATGTAATTCATGTGCACCTTGC AAAAGGAGATCTCGACATGCAAAAT	6638		
Db	330	TTTTACACCAACCAATGTAATTCATGTGCACCTTGC AAAAGGAGATCTCGACATGCAAAAT	271		
QY	6639	GTTCACGAGAACAAACCCAGCTTTTGTGCCAAGGTGACTGTAACTCAGATGGAAGTGG	6698		
Db	270	GTTCACGAGAACAAACCCAGCTTTTGTGCCAAGGTGACTGTAACTCAGATGGAAGTGG	211		
QY	6699	GCATTATATAAGGTTGGAGTGAAGAACATGCTGTATGTTTACTTAAACAGCCCTTTGAATT	6758		
Db	210	GCATTATATAAGGTTGGAGTGAAGAACATGCTGTATGTTTACTTAAACAGCCCTTTGAATT	151		
QY	6759	TAACAAAAACCTGGGAATCCATTAGGAACCGGATTCGATCATACCTGCAACATTAAGCTGAC	6818		
Db	150	TAACAAAAACCTGGGAATCCATTAGGAACCGGATTCGATCATACCTGCAACATTAAGCTGAC	91		
QY	6819	TGCTGAAATGTATTTTTTAGCTAATGAAAAGTGTTTGGACTAGTACTCTAAAAATGTTTC	6879		
Db	90	TGCTGAAATGTATTTTTTAGCTAATGAAAAGTGTTTGGACTAGTACTCTAAAAATGTTTC	31		
QY	6879	TAATGATAAAGTTTTTGAGTCAAAATAGAAA	6908		
Db	30	TAATGATAAAGTTTTTGAGTCAAAATAGAAA	1		

## RESULT 13

BG253502

## LOCUS DEFINITION

\*\*\*\*\*

ACCESSION

**VERSION**  
**KEYWORDS**

## NETWORK SOURCE

ORGANISM

## REFERENCE

## AUTHORS

TITLE  
JOURNAL.

COMMENT

[illegible]

RESULT 12	AW339499/c	570 bp	linear	EST 31-JAN-2000
LOCUS	AW339499		mRNA	
LOCUS	x291205.x1		NCI CGAP clone	IMAGE:2871608 3',
DEFINITION	Homo sapiens CDNA clone IMAGE:2871608 3', mRNA sequence.			
ACCESSION	AW339499			
VERSION	AW339499.1	GI:6836125		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
	1 (bases 1 to 570)			
REFERENCE	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .			
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
TITLE	Tumor Gene Index			

Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabp8-remail.nih.gov](mailto:cgabp8-remail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40UP from Gibco  
High quality sequence stop 471.

```

FEATURES
source
high quality
Location/Qualifiers
1..570
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2871608"
/tissue_type="carcinoid"
/lab_host="VDH108"
/clone_lib="NCI_CGAP_Lu24"
/note="Organ: lung; Vector: pYT3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized
library NCI CGAP Lu5 was prepared, and ss circles were

```

Qy 6270 GAAGGCAGTCTTT 6284  
||||| | | |  
Db 655 GAAGGCAGGCCCTT 669

RESULT 14  
AG152578/C

LOCUS	AG152578	659 bp	DNA	linear	GSS 09-JAN-2002
DEFINITION	pan troglodytes dna				

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
Ag152578	Ag152578.1	GI:16682256	GSS.	Pan troglodytes (chimpanzee)	1	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Toki, Y., Watanabe, H. and Sakaki, Y.	BAC end sequences of Library RPc1-43
				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.			

TIME  
JOURNAL  
BAC end sequences of Library RPCI-43  
Unpublished

FUJII  
 TITLE  
 DIRECT SUBMISSION  
 FUJIYAMA, A., HATTORI, M., TOYODA, A., TAYLOR, T. D., YADA, T.,  
 TOTOKI, Y., WATANABE, H. and SAKAKI, Y.  
 DIRECT SUBMISSION

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: [chimbess@gsc.riken.go.jp](mailto:chimbess@gsc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

end was generated during the R&D process and may have higher chance of clone tracking errors.  
PRIMERS

Sequencing: 10  
LIBRARY  
Vector : pACE3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI

FEATURES	
source	Location/Qualifiers
	1. .659

```

/organism="Pan. troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/col_xref="taxon:9598"

```

```

/sex="male"
/cell_type="lymphocytes"
/clone_id="RPC1-43 Chimpanzee Male BAC Library"

```

ORIGIN	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	7
--------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---

best Local Similarity 94.5%; Pred. No. 6.1e-48;  
Matches 598; Conservative 0; Mismatches 32;  
Indels 3; Gaps 2;

656 AACCCAGTTTGTCCCCAGGTGGTGTAACCTCAAAATGGGAAGTGGCTTATATAATAGG 597

Accession	Sequence	Length
6712	GTGTGGAGTGAAGAACATGCTGTAATGTTACTTAACAGCCCTTGGATTTTAAACAAAAAC-TG	6770
596	GTGTGGAGTAAAGAACATGCTGTAATGTTAGTACAGGCCCTTTGAATTTTAAACAAAACGTG	537

6771 GGAATCCATTAGGAAACCGAATTGCATCATCCTGAACATAAGCTGGACTGCTGAAATTGT 6830

6831 ATTTTAGCTAATCAAAAAGTGTTTGGACTAGTACTCTAAAAAGTCTCTAATGATAAGT 6890

6891 TTTGAGTCAAAATAGAAAAAGAAAAAATCTGCATTCAGGCCGAATTTTGTATATTTTA 6950

```

      Sequencing: -21M13
LIBRARY
    Vector      : pKS145
    R.Site 1   : SacI
    R.Site 2   : SacI
    Location/Qualifiers
        1..673
            /organism="Pan troglodytes"
            /mol type="genomic DNA"
            /db xref="taxon:9598"
            /clone="PTB-129N20.F"
            /sex="male"
            /cell type="lymphoblast"
            /clone lib="PTB Chimpanzee Male BAC Library"
238 a      127 c      120 g      188 t
BASE COUNT
ORIGIN

```

Search completed: February 1, 2004, 00:08:15  
Job time : 13523 secs

Search completed: February 1, 2004, 00:08:15  
Job time : 13523 secs